Biochemical and structural characterisation of a two-component signalling system downstream of bacteriophytochrome photoreceptor 1 in *Rhizobium* NT-26

Marta Wojnowska

Institute of Structural and Molecular Biology

University College London

Thesis submitted for the degree of Doctor of Philosophy

January 2013

Declaration

I, Marta Wojnowska, declare that the work presented in this thesis is my own. All the experiments were performed by me personally in the laboratory of Dr Snezana Djordjevic except for those listed below. Mass spectrometry experiments were performed in collaboration with Dr Konstantinos Thalassinos and his group; ESI MS and ion mobility data were acquired and analysed by Jun Yan, while LC-tandem MS data were acquired and processed by Adam Cryar. Collection of analytical ultracentrifugation data was conducted by Jayesh Gor in the laboratory of Prof. Steve Perkins. Electron microscopy studies were performed in the laboratory of Prof. Helen Saibil (Birkbeck College) with the help of her research group. PA synthesis was conducted in the laboratory of Dr Stefan Howorka (Department of Chemistry). Radioactive assays were performed in the laboratory of Prof. Finn Werner and the former laboratory of Dr Irina Tsaneva. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Prokaryotic signal transduction frequently involves two-component systems (TCSs), typically comprising a sensor histidine kinase (HK) and a response regulator (RR). Via a conserved phosphotransfer reaction, TCSs couple the detection of diverse stimuli with appropriate responses.

The initial aim of this project was to characterise the two-component "signalome" of an arsenite oxidiser, *Rhizobium* NT-26, in the context of the environmental niche, and compare it to "signalomes" of other bacterial species. A light-sensing HK, bacteriophytochrome photoreceptor 1 (BphP1), was thus identified. Previous studies indicated that BphP1 and the members of its gene cluster - two single-domain RRs and a hybrid RR/HK protein, ExsG - may constitute a TCS. Functional characterisation revealed that BphP1 initiates a branched signalling pathway; however, the mediated output could not be identified.

ExsG, a HWE-type HK, was shown to possess dual HK/RR activity and act as a novel type of signalling switch: phosphorylation of the N-terminal receiver domain downregulates the autokinase activity of the C-terminal kinase core. Ion mobility-mass spectrometry analysis indicated that while Asp62 phosphorylation stabilises the "closed" form of ExsG, nucleotide binding stabilises the "open" conformation. These observations led to a model in which phosphorylation of the receiver domain precludes ATP binding and thus inhibits autokinase activity. Furthermore, ExsG was demonstrated to hexamerise via the HWE core, which makes it the first non-dimeric HK. Notably, however, the basic unit of the hexameric assembly is a homodimer, and the HWE core shares homology with canonical HK cores.

The results presented herein broaden the current knowledge on TCSs and identify previously unreported mechanisms involved in two-component signal transduction. The members of the BphP1 signalling cascade enrich the pool of modular communication units that can be exploited in engineering artificial signalling networks, biosensors and microorganisms with novel functionalities.

Acknowledgements

First of all, I would like to thank my supervisor, Dr Snezana Djordjevic, for the patience, understanding, advisory - not only in the context of research - as well as for the efforts to turn me into an open-minded, knowledgeable scientist. Her encouragement and the "I won't stop you from doing this"-attitude helped me gain a considerable dose of research experience.

Without the continuous support from my family and friends I would have never got here. I cannot fully express my gratitude to my aunt and uncle for everything they have offered and provided me with especially in the last few months. I am indebted to Gosia for keeping my spirits up and and her unselfish attitude, and to Sophie "Zabinski" for all those pints and discussions, as well as for proofreading my thesis. I would like to thank my family for their never-ending support and all of my friends who made me feel a valuable person, provided encouragement and helped me retain sanity, particularly the Dziki (Kinga, Basia, Justin), Bea, 168 North Gower Street, Irene and the 6th floor lab members.

I need to extend my gratitude to Dr Thalassinos and his group - Jun, Ganesh and Adam - for a fruitful collaboration that allowed us to perform various exciting experiments. I am also thankful to Prof. Saibil and the EM lab members for assistance with electron microscopy, Prof. Perkins and Jayesh for help with analytical ultracentrifugation, as well as Dr Howorka and Dr Burns for the assistance with phosphoramidate synthesis. Furthermore, I would like to thank Wellcome Trust for funding and the ISMB for providing a great working atmosphere.

Finally, I would like to thank Tamas for supporting me in all possible ways, particularly when this thesis was being written - ILYSFM.

Contents

List of Figures

List of Tables

List of abbreviations

β-ME	β-mercaptoethanol
AAA	ATPases associated with various domain activities
ABD	ATP-binding domain
AC	adenylyl cyclase
AMP-PNP	5'-adenylyl-β-γ-imidodiphosphate
AUC	analytical ultracentrifugation
BLAST	basic local alignment search tool
bp	base pairs
BphP	bacteriophytochrome photoreceptor
BphP1-HK	BphP1 histidine kinase core
CA	catalytic and ATP-binding
CATH	class, architecture, topology, homology
(k)Da	(kilo)Dalton
DHp	dimerisation and histidine phosphotransfer
DMSO	dimethyl sulfoxide
dNTPs	deoxy-nucleotide triphosphates
DTT	dithiothreitol
EAL	containing conserved E, A, L residues
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ESI	electrospray ionisation
ExsG-HK	ExsG histidine kinase core
ExsG-REC	ExsG receiver domain
GAF	cGMP-specific phosphodiesterases, adenylyl cyclases, FhlA
GGDEF	containing conserved G, G, D, E, F residues
GHKL	gyrases, Hsp90, histidine kinases, MutL
HAMP	histidine kinases, adenylyl cyclases, methyl binding proteins, phosphatases
HCD	histidine-containing domain
HK	histidine kinase
Hpt	histidine phosphotransfer
HWE	containing conserved H, W, E residues
IM	ion mobility
IPTG	isopropyl-β-D-thiogalactopyranoside

ITC	isothermal titration calorimetry
LB	lysogeny broth
MALDI	matrix-assisted laser desorption/ionisation
MCP	methyl-accepting chemotaxis receptor protein
MES	2-(N-morpholino)-ethanesulfonic acid
MS	mass spectrometry
MWCO	molecular weight cut-off
NiNTA	Ni^{2+}-nitrilo-triacetic acid
PA	phosphoramidate
PAGE	polyacrylamide gel electrophoresis
PAS	period, aryl-hydrocarbon receptor nuclear translocator and single-minded
PCR	polymerase chain reaction
PDB	Protein Data Bank
PP2C	protein phosphatase 2C
PSL	photostimulated luminescence
PTA	phosphotungstic acid
RR	response regulator
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
SE	sedimentation equilibrium
SMART	simple modular architecture research tool
STYK	serine/threonine/tyrosine protein kinase
SV	sedimentation velocity
TAE	Tris, acetic acid, EDTA
TCS	two-component system
TEV	tobacco etch virus
ToF	time-of-flight
WT	wild-type

Chapter 1

Introduction

The ability to respond to environmental changes is vital for all organisms. Stimulus detection needs to be coupled with an appropriate adaptation and requires signalling cascades of varying degree of complexity. Signal transduction in prokaryotes tends to be less intricate than in eukaryotes but bacteria also need to respond to diverse stimuli - nutrients, toxic compounds, light, temperature, chemical gradients, pH changes, membrane fluidity, and physical forces. A common form of microbial signal transduction is a two-component system (TCS) which relies on phosphotransfer between two communicating modules, a histidine kinase core and a receiver domain [1, 2, 3, 4]. These two domains involved in phosphotransfer can be coupled with various input and output domains which allows the assembly of diverse signalling pathways. The high level of conservation of the phosphotransfer domains demonstrates how a successful signalling mechanism has been recycled by evolution.

Absence of TCSs from higher eukaryotes and their frequent involvement in bacterial virulence make them promising antimicrobial targets [5, 3]; their intrinsic modularity can be successfully exploited in designing new signalling proteins and circuits [6, 7, 8], as well as microbes with programmable properties [9, 10]. Given the potential of exploiting TCSs in synthetic biology and medicine, it is crucial to identify all the stimuli sensed and responses triggered, as well as to fully understand the mechanisms involved in this form of signal transduction.

1.1 Two-component signal transduction

Two-component signalling involves two basic units which can be arranged into pathways of varying complexity, but communication always involves shuttling a phosphoryl group from histidine onto aspartate (Fig. 1.1A) [3, 11, 12, 2]. In a canonical TCS the sensor

protein, histidine kinase (HK), detects the stimulus via its N-terminal sensing region, and the signal is relayed to the catalytic core, which acts as the transmitter. All classical HKs characterised so far form bilaterally symmetrical homodimers via their kinase cores and signal transmission from the sensing domain to the core usually leads to exposure of the active site histidine so that it is positioned closer to the ATP-binding domain [13]. When active, the protein autophosphorylates upon a conserved histidine residue and the phosphotransfer onto the invariant aspartate on the receiver domain of a cognate response regulator (RR) occurs [3, 4]. Phosphorylation of the RR receiver domain triggers conformational alterations which are coupled with activation of the C-terminal effector domain. RR phosphorylation levels are often negatively regulated by phosphatase activity or RR autodephosphorylation; the response triggered by each TCS depends on the proportion of phosphorylated RRs.

The number of proteins belonging to a TCS varies - canonical pathways involve two proteins, one with sensor and transmitter functions and the other acting as the receiver/effector. However, HK core and receiver domain have been frequently exploited by evolution as building blocks for constructing larger multistep signalling pathways, potentially providing more phosphorylation checkpoints or space for additional regulatory mechanisms to develop [14]. Moreover, gene duplications, fusions and other evolutionary events also shaped the architectures of some TCSs; there are notable cases where HK and RR are fused together [15] or of proteins with unusual domain arrangements, such as hybrid RRs [16] (see Section 1.3.2).

The more complex variant of two-component signal transduction, phosphorelay, uses several phosphotransfer steps (His-Asp-His-Asp) and usually entails a hybrid HK protein, i.e. a HK with a receiver domain appended at the carboxyl terminus; the receiver domain may also exist as a stand-alone protein (Fig. 1.1B) [3, 4]. The other essential signalling module in phosphorelay is a histidine phosphotransfer (Hpt) domain which can exist as a separate protein or as a further C-terminal addition to the hybrid HK. After the first, frequently intramolecular, phosphotransfer, the phosphoryl group is shuttled from the receiver domain aspartate onto a histidine residue of the Hpt domain, and further onto the aspartate of another RR, which usually contains an effector domain.

There are also intrinsically branched systems, following a one-to-many or many-to-one scenario [17], where one HK phosphorylates several RRs or where signalling from several HKs converges at a single RR. Furthermore, hybrid signalling systems exist where a TCS is directly or indirectly coupled with another form of signal transduction, such as protein serine/threonine/tyrosine kinases [3, 18].

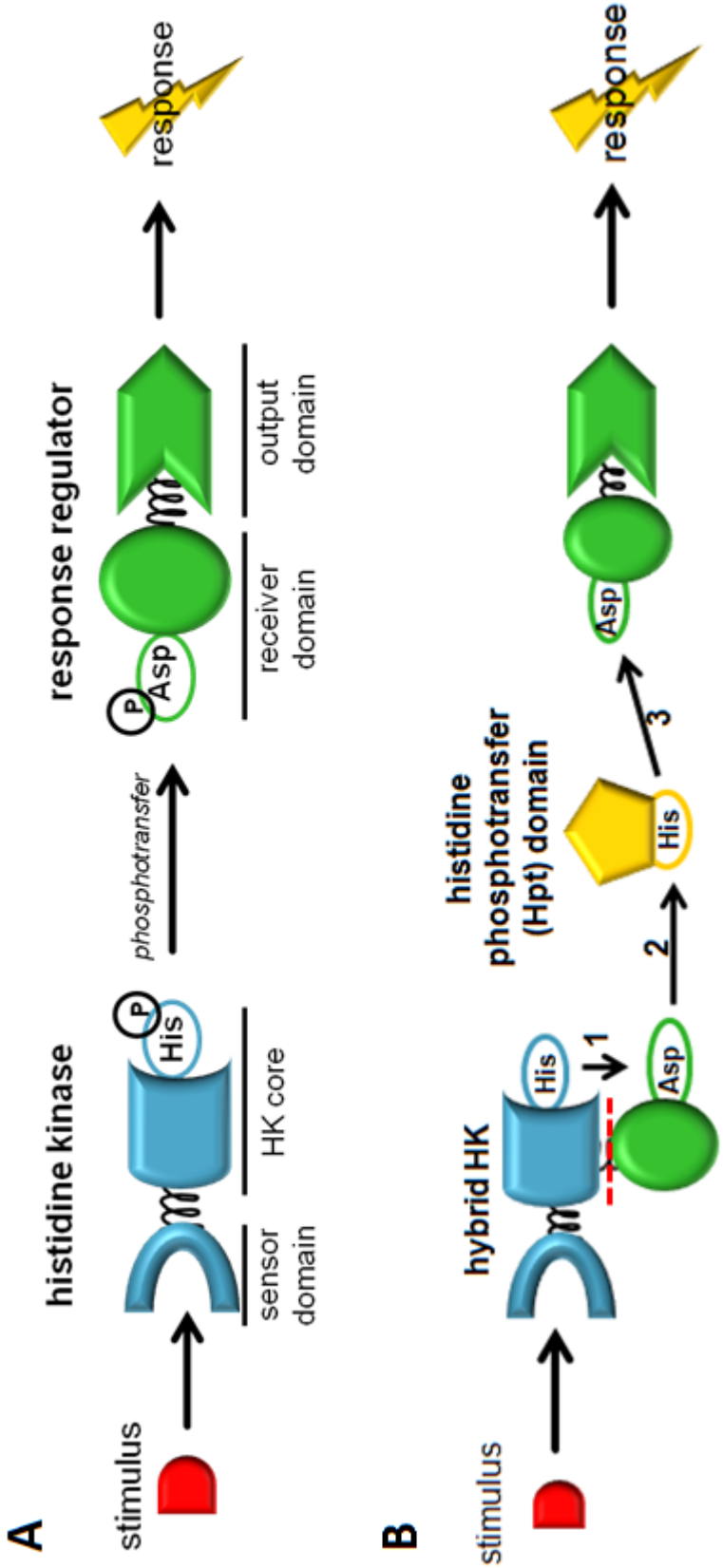


Figure 1.1: Schematic representation of a canonical TCS (A) and phosphorelay (B). Black numbered arrows denote subsequent phosphotransfer steps. Red dashed line indicates that the intermediary receiver domain in phosphorelay is usually appended to the sensor kinase, but may exist as a single domain protein. Attachment of the Hpt domain to the C-terminus of a hybrid HK is possible but not shown here.

1.1.1 Stimulus detection by sensor domains

As previously mentioned, intrinsic modularity of TCSs allows diverse combinations of sensing and transmitter domains [3, 4, 19, 20, 21]. While the HK cores exhibit high sequence and fold conservation, the sensory regions are the primary source of variation within the HK superfamily. Stimulus detection is normally confined to the N-terminal portion of the protein which can comprise a single domain as well as an array of domains. Diverse sensory domains can be associated with HK cores, and many of them have not yet been characterised in terms of structure and function. However, there are also some domains which occur frequently, such as PAS (period, aryl-hydrocarbon receptor nuclear translocator and single-minded) and GAF (cGMP-specific phosphodiesterases, adenylyl cyclase, FhlA) domains found in cytoplasmic HKs [22, 23]. They adopt a similar fold involving a central antiparallel β-sheet flanked by helices on either side, with the difference only in the number of β-strands [24]. These cytoplasmic domains are highly variable in terms of sequence and function and may be associated with diverse cofactors. Numerous membrane-associated HKs utilise periplasmic PAS-like domains for sensing extracellular signals; such sensing domains are related to the PAS fold via the central β-sheet topology but display different arrangement of the helical regions [25, 4]. There are also various transmembrane and periplasmic domains that are commonly associated with HKs, for instance the periplasmic-sensing CHASE (cyclases/histidine kinases associated sensory extracellular) family of domains and membrane-integral MHYT (conserved Met, His, Tyr and Thr residues) domains [19]. These domains can be involved in sensing diverse stimuli, such as small molecules, turgor or temperature.

The activity of the HK core is usually regulated by the N-terminal sensory (transmembrane or cytoplasmic) domain or an array of domains, but some HKs require additional proteins for sensing. In transmembrane HKs, the stimulus - for example, the binding of a ligand or a change in membrane fluidity - affects the conformation of the sensing domain, which in turn leads to the movement of the linker region between the transmembrane or extracellular domains and the cytoplasmic HK core. These conformational perturbations induce, depending on the system, the active or inactive state of the HK core.

1.1.1.1 Stimulus detection mechanisms

Stimulus detection can be confined to the N-terminal part of the HK or, occasionally, to separate proteins which modulate HK activity either by interaction with the N-

terminal region or with the HK core itself [21, 19]. The use of accessory proteins means that the HK detects the stimulus indirectly; the best studied examples are LuxQ whose periplasmic domain is in complex with the periplasmic binding protein LuxP, and cytoplasmic NtrB which is regulated by the binding of protein PII to its ATP-binding domain. LuxP binds the autoinducer and the conformational changes within LuxP affect the state of LuxQ. PII protein under nitrogen limitation conditions is present in a uridylated form that cannot bind to NtrB thus allowing autokinase activity.

Sensor domains can be classified into three main groups - cytoplasmic, transmembrane and periplasmic, with the latter being the most common type associated with HKs [19, 25]. Periplasmic domains frequently bind chemical compounds; many of them assume a PAS-like or an entirely helical fold (a four-helix bundle), while some share a common topology with periplasmic binding proteins [25]. Both PAS-like and all-alpha types of fold may be associated with various domain insertions or extensions. Sensing domains that involve multiple membrane-spanning regions, without extruding periplasmic regions, are thought to sense changes associated primarily with the membrane; however, some of them have also been implicated in forming protein-protein interactions or binding various natural ligands [25], as well as antibiotics [19, 21], via transmembrane helices or the extracytoplasmic loops.

Cytoplasmic-sensing HKs exist both as membrane-embedded as well as fully soluble proteins [19, 25]. They can be involved in detection of diffusible stimuli, such as oxygen, and changes in developmental or metabolic state. As previously mentioned, PAS and GAF domains are commonly found in cytoplasmic HKs and can be involved in detection of both chemical and physical stimuli. Soluble HKs display diverse domain architectures and may incorporate a series of domains, either of the same type or in combination with other domains. HKs with multiple sensory domains are likely to monitor several signals at the same time. The possibility of simultaneous sensing of multiple signals has been postulated to be the primary reason for the existence of cytoplasmic HKs; one-component systems, where the sensor domain is directly attached to the effector domain, do not normally contain more than one sensing domain [26].

1.1.1.2 Signal propagation to the catalytic core

Detection of a stimulus triggers conformational changes which regulate the activity of the kinase core. However, given the diversity of sensor domains and the complexity of certain sensing systems, the mode of signal transmission may require different structural features and mechanisms.

Signal transduction across the membrane still remains largely a mystery, although

in some cases it appears to entail piston-like movement of the helices towards and across the membrane [4]. Many HKs contain a HAMP (histidine kinases, adenylyl cyclases, methyl binding proteins and phosphatases) domain which links the transmembrane region with the HK core. HAMP domains dimerise to form a four-helix bundle and it has been shown that transmembrane signal transmission involves rotation (cogwheeling) of the helical bundle with each helix rotating in the opposite direction to the adjacent one; this in turn affects the relative positioning of the downstream HK core subunits [27]. However, whether this mechanism applies to all canonical HKs is unknown. In addition, a coiled-coil structure termed the "signalling helix" has been identified in numerous signalling proteins, including HKs, and its conservation may imply a common signal transduction mechanism [28, 4]. At least for some HKs, the change from inactive to active state may require partial unwinding of the helix containing the phosphorylatable histidine residue [29].

Another way of regulating HK activity relies on the symmetry of the HK homodimer [30]. As mentioned before, the sensory domain of LuxQ is permanently associated with LuxP. In the absence of autoinducer, the cytoplasmic portion of LuxQ homodimer forms a symmetric four-helix bundle. Autoinducer binding to LuxP causes conformational changes which induce an asymmetric association of the LuxQ kinase cores thus inhibiting autokinase activity.

In cytoplasmic HKs, such as FixL, direct binding of ligands can affect the conformation of the protein, altering interdomain contacts and thus transmitting the signal to the catalytic core [21]. The sensory region of FixL and many other HKs comprises a PAS domain, which is a versatile sensing module with a common topology. In addition, several transmembrane HKs also utilise a cytoplasmic PAS domain; the frequent occurrence of this type of fold implies that many HKs comprising a PAS domain would share a similar signal propagation mechanism. However, in cases where one HK possesses multiple sensory domains, signal transmission to the catalytic core may involve different mechanisms.

The activity of HKs can also be regulated by the formation of disulfide bonds within the sensory domain, as in the case of the redox sensors ArcB and RegB [21]. ArcB PAS domain can enforce dimerisation of the protein via formation of disulfide bridges, and this is dependent on the redox state of the cell.

1.1.2 Signalling mechanism

HK core comprises two domains which are schematically depicted in Figure 1.2. [3, 4]. Dimerisation and histidine phosphotransfer (DHp) domain harbours the phosphoryla-

table histidine and is linked to the sensory or transmembrane region via a signalling linker, such as HAMP domain or signalling helix, or through a cytoplasmic sensory region, e.g. a PAS domain. Catalytic and ATP-binding (CA) domains contain several highly conserved amino acid motifs crucial for ATP binding and hydrolysis. Interestingly, even when present as a separate protein, the CA domain is capable of recognising and phosphorylating the histidine located within the DHp domain [31]. Several HKs have been shown to autophosphorylate in *trans*, whereby the CA domain of one subunit phosphorylates the histidine residue located on the other subunit, but in some HKs autophosphorylation can also occur intramolecularly (in *cis*) [32, 4]. The CA domain is mobile due to the presence of a flexible linker between the two domains [13] and autophosphorylation is thought to require changes in positioning of the CA domain with respect to the DHp domain [33, 13]. The currently accepted model for autokinase activity regulation, which is postulated to involve CA domain sequestration and cogwheel-like rotation of DHp helices [13], is shown in Figure 1.3.

The equilibrium of the autophosphorylation reaction favours the unphosphorylated protein, hence only a small proportion of HKs will exist in the phosphorylated state within the cell [3]. Both phosphohistidine and phosphoaspartate are unstable, which facilitates signalling and allows rapid reversion to the unphosphorylated state [3, 4]. Divalent metal ions are required for autophosphorylation and phosphotransfer, as well as for dephosphorylation reactions.

1.1.2.1 HK autophosphorylation mechanism

As described above, there are various ways of transmitting the signal from sensor regions to the homodimeric catalytic cores. There also seem to be several mechanisms of regulating autophosphorylation. Being directly linked to the sensory domain, the DHp domain is likely to be directly affected by the conformational changes upstream. HAMP domains have been demonstrated to undergo cogwheeling [27] and these rotational changes may be transmitted further onto the DHp helices, leading to domain rearrangements and potentially reducing the distance between the ATP and phosphorylatable histidine [13].

When the two domains of the HK catalytic core assume the active conformation, ATP can be used to phosphorylate the conserved histidine residue within the DHp domain. The available structures of HKs do not seem to have captured the active state, but both inter- and intra-molecular phosphorylation may be possible: Casino and co-workers showed that in the case of HK853 the CA domain phosphorylates His within the DHp domain located on the same polypeptide [32], but previous studies

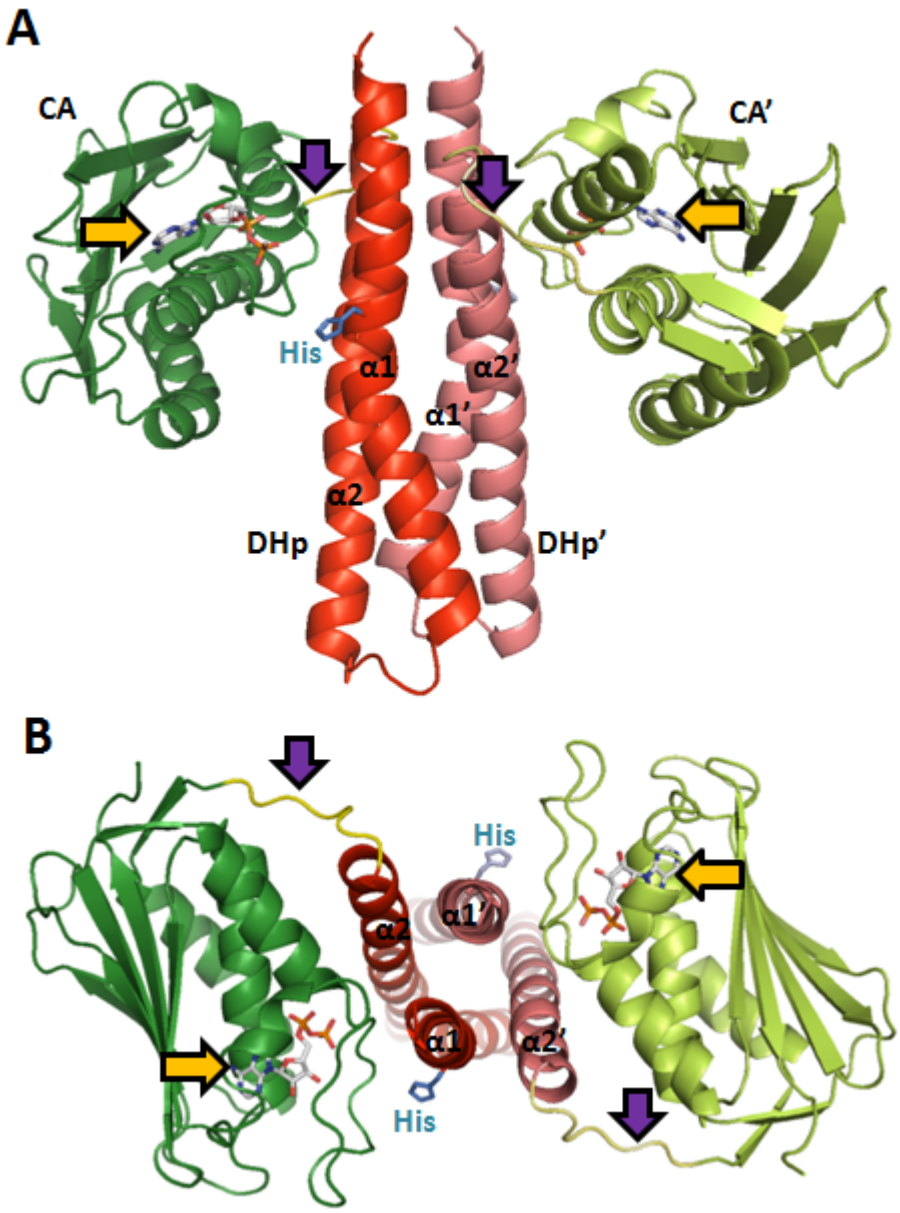


Figure 1.2: Structure of *Thermatoga maritima* HK853. The molecular orientation in A was rotated by 90° to generate the view in B, down the axis of the four-helix bundle. CA domains are shown in green, DHp in red, the flexible linker (indicated by purple arrows) is in yellow. Phosphorylatable histidine residues are shown as blue sticks. Orange arrows indicate the nucleotide (ADPβN, the product of AMP-PNP hydrolysis) bound to the CA domain. PDB ID 3DGE [32].

reported *trans*-phosphorylation between the two dimerised subunits [34, 35]. This difference in autophosphorylation mode has been attributed to the dissimilar arrangement of the four-helix bundle in HK853-like proteins and in HKs like EnvZ, exhibiting *trans*-phosphorylation [32]. It seems that both modes occur and they share the same mechanism, except that the phosphotransfer occurs intra- or intermolecularly [13].

DHp has recently been postulated to undergo partial unwinding upon the reception of a stimulus by the sensory domain; this would break DHp-CA interactions stabilising the inactive conformation, and new contacts can be formed which allow the transfer of phosphate from CA to the histidine residue within the DHp domain [29]. Structures of a HAMP domain fused to DHp domain of EnvZ also support the idea that changes within the region directly upstream of the DHp domain abolish its ability to interact with the CA domain, permitting the two domains to reorient in a fashion that suits autophosphorylation [33]. In general, it appears that conformational abberations within the DHp domain, or the region at the junction of the DHp domain and the signalling linker, are required to free the CA domain which in the inactive form is sequestered (Fig. 1.3) [13]. Sequestration of the catalytic domain is essential to separate the different activities of the HK: autophosphorylation, phosphotransfer, and - in some cases - RR dephosphorylation. Contacts between the CA domain and the RR receiver domain, observed in the structure reported by Casino and co-workers, may be necessary for providing additional binding interface for the RR during phosphotransfer or dephosphorylation [32]. In that structure, HK853 seems to be in the phosphotransfer or phosphatase state and the CA domain is positioned away from the phosphorylatable histidine, which allows the RR to bind to the site of phosphorylation [32, 13].

The crystal structure of HK853 in complex with its cognate RR468 indicates that other protein-protein interactions may also influence the ability to autophosphorylate [32, 36]. It appears that the RR is in contact with the flexible ATP-lid of the CA domain, which traps the ADP molecule, thus precluding further autophosphorylation. The ThkA/TrrA complex structure reported by Yamada and coworkers presents yet another putative regulatory mechanism: the interaction between the ThkA PAS and CA domains imposes separation of the CA domain from the phosphorylatable histidine and seems to be essential for the phosphatase activity [37]. In addition, the interaction between DHp and CA domains can also be regulated by other proteins; KinA and KinB autophosphorylation is regulated by two different "antikinase" proteins which are able to preclude effective DHp-CA domain interaction by binding to the DHp domains [38, 39]. NtrB catalytic activity is influenced by the binding of PII protein, as mentioned before [19, 21]. In the case of DesK, a thermosensor HK, both phosphorylation-induced helical

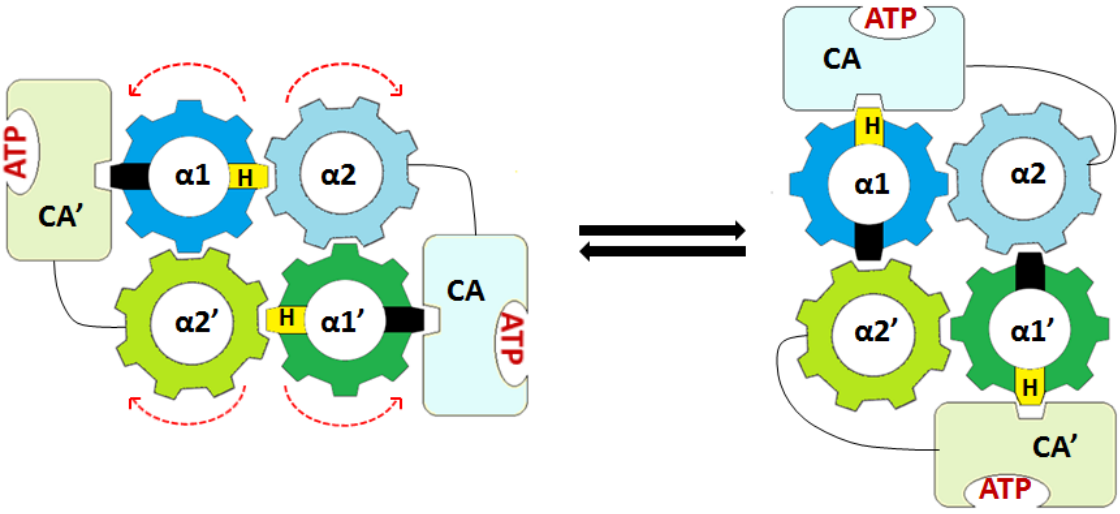


Figure 1.3: Model of autokinase activity regulation involving cogwheeling and CA domain sequestration. The pairs of helices of DHp domains are represented by cogwheels and viewed from the top as in Fig. 1.2B. When a portion of DHp domain responsible for sequestration of the CA domain is exposed (black notch), the kinase activity is inhibited while phosphotransferase and phosphatase activities are switched on (left). Stimulus-dependent rotation of the helices, indicated by red dashed arrows, is transmitted from HAMP domain or signalling helix to the helical bundle. This rearrangement hides the sequestering patches of the DHp domains thus releasing the CA domains (right). Simultaneously, phosphorylatable histidine residues (yellow notches) become exposed and oriented in a way that allows autophosphorylation. Based on Stewart 2010 [13].

kink of one of the DHp domains and signalling from the upstream transmembrane region via helical rotations may regulate the autokinase activity [40]. The bending of helix α1, observed in DesK and KinA, seems to involve isomerisation of a conserved proline residue, implying that this amino acid may play a role in regulating autophosphorylation [13].

1.1.2.2 Phosphotransfer reaction

Phosphotransfer is catalysed by the receiver domain of the RR, which is evidenced by the fact that the receiver domains can be phosphorylated by phosphohistidine-containing HK fragments and small-molecule phosphate donors, such as phosphoramidate or acetyl phosphate [41, 3]. Several residues on both HK and RR are responsible for phosphotransfer specificity [32, 37]. The divalent cation, usually magnesium ion, is coordinated by the two receiver domain residues, which are commonly an aspartate or a glutamate. Phosphotransfer inolves a nucleophilic attack by the carboxyl oxygen of the specific aspartate residue of the receiver domain, presumably followed by formation

of a pentavalent phosphorus intermediate.

1.1.2.3 Phosphohistidine and phosphoaspartate

Serine/threonine/tyrosine kinases form phosphoesters while HKs produce a phosphoramidate bond. This high energy N-P bond is unstable primarily due to the poor overlap between the nitrogen orbital containing the lone electron pair and the phosphorus d-orbital which form the π-bond; as a consequence electron delocalisation, which stabilises amide bonds, has insignificant contribution to the stability of the phosphoramidate bond [42]. In the case of phosphohistidine, the phosphoramidate nitrogen is non-basic as its lone pair is delocalised over the imidazole ring. Acid lability of phosphohistidines is linked to protonation of the non-phosphorylated imidazole nitrogen, which reduces electron density at the phosphoramidate nitrogen and allows P-N bond hydrolysis. The instability of the phosphoramidate bond is a crucial feature of TCSs as it ensures rapid transfer of phosphoryl groups onto RRs [42, 3].

Phosphoaspartate is generally unstable, but its half-life is correlated to the nature of the response triggered by the respective TCS [3, 43]. Half-lives can vary from less than a second to weeks [43]. Processes occurring within narrow time scales, like chemotaxis, require rapid dephosphorylation of RRs in order to finely tune the response. On the other hand, signal transduction which leads to gene expression may involve a more stable phosphoaspartate. The receiver domain is capable of autodephosphorylation and several residues have been implicated in enhancing or reducing Asp∼P stability [44, 45]. The identity of these residues within the receiver domain sequence can be used to estimate the stability of phosphoaspartate and may allow classification of the type of response triggered by the TCS system employing such receiver domain.

Detection of phosphoaspartate is difficult to achieve due to its inherent temporal and chemical instability. However, recently a new electrophoresis method employing Phos-tag^{TM} acrylamide enabled separation of phosphorylated protein and identification of (Asp∼P)-proteins [46, 4].

1.1.2.4 Negative regulation

RRs function as binary switches and it is the proportion of phosphorylated molecules that is crucial for evoking the appropriate adaptation [3]. As mentioned above, some systems require rapid reversion to the inactive state after triggering the response to enable highly sensitive and robust responses. In addition, all TCSs seem to have intrinsic mechanisms regulating RR phosphorylation levels which ensure that RR∼P levels

increase only when the cognate HK is phosphorylated, thus eliminating the effects of potential cross-talk [17, 47].

Both RR autodephosphorylation and phosphatase activity may be employed by TCSs to regulate RR phosphorylation levels. Receiver domains are capable of hydrolysing the Asp∼P bond but the time scales differ significantly depending on the type of response triggered [3, 4, 43]. Phosphatase activity may be confined to the cognate HK or an auxiliary protein as in the case of chemotaxis system where CheZ and CheX phosphatases are responsible for dephosphorylating the RR CheY [3, 4]. Chemotaxis system couples the detection of attractants and repellents with changes in flagellar motor rotation, thus affecting the motility of the bacterium. Many HKs have been reported to be bifunctional - when unphosphorylated, they dephosphorylate their cognate RRs. Some chemotaxis systems involve a CheY-like protein acting as a phosphate sink, receiving the phosphate flowing from CheY∼P through CheA. Moreover, various two-component "connector" proteins exist which via interactions with HKs or RRs regulate phosphorylation levels of RRs [48].

Phosphatase activity of bifunctional HKs can be induced by interdomain or protein-protein interactions. In the case of soluble HK NtrB the binding of another protein, PII, controls the switch between autokinase and phosphatase activity [21, 19]. The structure of the ThkA-TrrA complex suggests that the PAS domain of ThkA may interact with TrrA and ThkA catalytic core in a way that promotes phosphatase activity [37]. The mechanism of dephosphorylation of RRs by cognate HKs is still not fully understood, but it does not seem to involve the reverse of phosphotransfer reaction and the active site histidine residue is not essential [49]. However, the DHp domain is usually sufficient for dephosphorylating cognate RRs, even though additional interdomain or protein-protein interactions may stimulate the phosphatase activity. ADP has been shown to enhance phosphatase activity [50]. In the case of phosphorelays which employ multistep phosphotransfers (His-Asp-His-Asp) dephosphorylation typically involves reverse phosphotransfer via the mediating histidine phosphotransfer (Hpt) domain.

1.1.3 Responses mediated by RRs

It is assumed that RRs exist in an equilibrium between active and inactive conformations [4], although intermediate conformations, that may be of physiological relevance, have also been reported [51, 52]. Phosphorylation of the aspartate residue shifts the equilibrium towards the active state; formation of the high-energy phosphoaspartate bond is thought to provide the energy for the long-range conformational changes within the receiver domain [3, 4]. Structures of phosphorylated and unphosphorylated

RRs revealed varying levels of backbone displacement but in all cases the movements affected the α4/β5/α5 region, implicated in mediating protein-protein interactions (see Section 1.2.3) [53, 3, 4]. Phosphorylation leads to activation or inhibition of associated effector domains and can affect RR oligomerisation state.

The majority (63%) of RRs contain DNA-binding domains and induce changes in gene expression [4]. Others are involved in chemotaxis and phosphorelays or possess effector domains with enzymatic activities. Some output regions can also mediate protein-protein interactions. Stand-alone receiver domain proteins are common in bacterial species and usually act in chemotaxis pathways, regulating the direction of flagellar rotation, or participate in multistep phosphotransfer systems [54]. Some single-domain RRs have also been shown to act as phosphate sinks or as allosteric regulators.

TCS signalling may involve signal convergence - several RRs can, for instance, induce expression of overlapping sets of genes [17]. In addition, some RRs regulate expression of their own genes or of genes encoding the entire TCS to which they belong [14].

1.1.3.1 Phosphorylation-induced changes in oligomeric state of RRs

Depending on the system, RRs can contain an appended effector domain or exist as stand-alone receiver domain proteins. The activity of associated output domains are regulated in a phosphorylation-dependent manner; it has been demonstrated that in most cases the α4/β5/α5 portion of the regulatory domain is involved in forming interdomain interactions [4, 55]. Conformational changes induced upon phosphorylation frequently affect the oligomerisation state: RRs with DNA-binding domains usually homodimerise in order to bind to DNA motifs, or oligomerise forming a ring structure when an AAA+ ATPase domain is present between the receiver and the DNA-binding domain [55]. Such changes may involve breaking up a pre-formed homodimer and the formation of new heterodomain interactions, as observed in NtrC proteins. Activation or inhibition of associated output domains may also involve allosteric regulation whereby changes in the interdomain contacts affect the conformation of the effector region [4, 55].

In contrast, CheY-like RRs are monomeric and bind to the protein FliM at the base of the flagellar motor via their α4/β5/α5 region. Rcp-type RRs, which are phosphorylated by cyanobacterial phytochromes, exist as homodimers or dimerise in a phosphorylation-dependent manner, but their downstream targets remain unknown [56, 57].

1.1.3.2 Types of responses

The most well studied systems are those regulating chemotaxis and several TCSs which induce gene expression, like EnvZ-OmpR or the *Bacillus subtilis* sporulation system. Chemotaxis system switches the direction of flagellar rotation in response to changes in attractant or repellent concentration in the environment. EnvZ senses changes in osmolarity and evokes the expression of porin genes by regulating the levels of OmpR phosphorylation. Sporulation control in *Bacillus subtilis* involves a more complex form of two-component signal transduction which ends with Spo0A, a RR which controls expression of proteins essential for entering the sporulation phase [3, 58]. There are also TCSs which are coupled with an enzymatic activity or, as in the case of eukaryotic TCSs, constitute a part of a larger signalling cascade involving other types of signal transduction [3, 18].

1.2 TCS signalling modules

Phosphotransfer modules utilised by TCSs are usually highly conserved and contain numerous signature motifs and invariant residues. HK catalytic cores are unrelated to eukaryotic Ser/Thr kinases, but instead belong to the GHKL ATPase superfamily [59, 3]. A non-canonical HWE HK core also exists that possesses an atypical ATPase region [60] but prior to this study it was not known if it is a member of the GHKL superfamily.

1.2.1 Histidine kinase

1.2.1.1 Canonical histidine kinase core

HK cores comprise two domains which exhibit different levels of conservation, but their overall topologies and functionalities are shared by all prototypical members of the HK family [4]. CA domain is located C-terminal to the DHp domain and its positioning with respect to the other domain has been shown to relate to specific functions of the HK. As previously described (Section 1.1.2.1), the distance between the CA domain and the active site histidine varies depending on the activity mode. DHp domains dimerise to form a helical bundle [61, 62, 32, 37, 40] while stand-alone CA domains have been shown to be monomeric [31, 63]. In the context of the full-length protein, the CA domain is linked to the preceding DHp domain through a flexible linker that allows for the plasticity of interactions between the two domains [13].

The DHp domain, termed HisKA in SMART [64] and Pfam [65] databases, encompasses a signature amino acid motif, termed the H-box, containing the phosphorylatable histidine residue and forms the primary interaction interface with the cognate RR. In addition, this domain is responsible for homodimerisation and frequently carries phosphatase activity. It is composed of two helices which, together with the DHp domain of another subunit, form a four-helix bundle (see Fig. 1.2). Active site histidine residue is located within α1 which is directly linked to the transmembrane or sensory domain; hence the His-containing region is likely to be directly affected by conformational changes linked to stimulus detection [13]. Homodimerisation of HKs is directed by interaction specificity - it seems that throughout evolution DHp domains of different HKs have diverged to ensure effective interactions [66]. Phosphotransfer to cognate RR also requires high fidelity of interaction which in at least some HKs is conferred to several residues and the inter-helix linker of the DHp domain [67]. As phosphotransfer specificity determinants differ between HKs, DHp domains exhibit lower level of conservation than the CA domains [4].

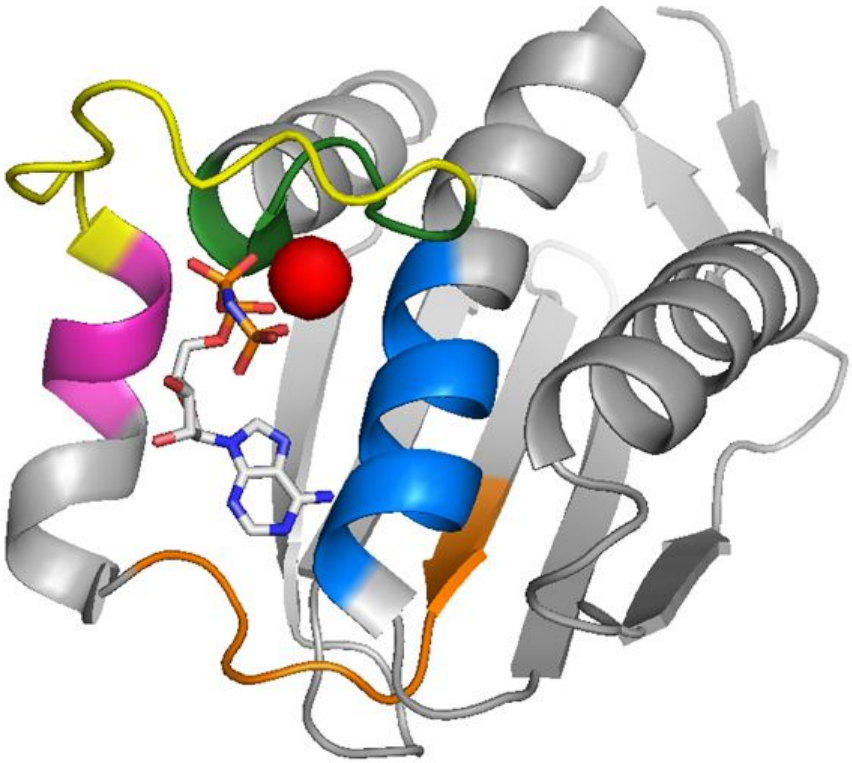


Figure 1.4: Structure of the nucleotide-bound CA domain of PhoQ. The four boxes containing the conserved amino acids are shown: N – blue, G1 – orange, F – magenta, G2 – green. ATP lid is in yellow. The magnesium ion is shown as a red sphere and AMP-PNP, an ATP analogue, is shown as sticks. PDB ID 1ID0 [68].

CA domains (HATPase_c in Pfam and SMART) belong to the large GHKL superfamily of ATPases and their mixed α/β sandwich (Bergerat) fold is highly conserved across the HK family, with several signature motifs defining the ATP-binding pocket (Fig. 1.4) [59, 4]. Bergerat fold employs three helices and five strands; the conserved residues coordinating the nucleotide are located within the loops connecting the structural elements [59]. The signature motifs - referred to as N, G1, F and G2 boxes - may exhibit different spacing across HKs but are normally arranged in this order [3, 59, 4]. Magnesium ion, coordinated by the conserved N-box asparagine, connects all ATP phosphate groups to the protein via solvent-mediated hydrogen bonding [59]. An aspartate within the G1 box forms a hydrogen bond with the N6 atom of the adenine moiety and the backbone atoms of G2 box glycines interact with ATP phosphates. The two glycine-rich G boxes provide flexibility for the ATP lid located between them while F-box, forming a part of the ATP lid, harbours residues which play a role in anchoring the flexible lid in absence of the nuclotide [62]. It appears that the ATP lid not only secures bound ATP but also locks in ADP, which is essential for phosphatase activity and for prevention of futile ATPase activity when the phosphatase mode is on [50]. The contacts between ATP lid of HK853 and its cognate RR provide further evidence for the involvement of this structure in promoting phosphatase activity [32].

There are also non-canonical domain arrangements as in the case of CheA where two different domains are responsible for dimerisation and autophosphorylation. CheA is a HK whose autophosphorylation is regulated by chemoreceptors. The protein is composed of five distinct domains and the phosphorylatable histidine is located within the N-terminal Hpt domain, at a considerable distance in terms of sequence from the catalytic domain [69, 70]. The CA domain is located directly next to a dimerisation domain, which is homologous with the DHp domain of classical HKs but lacks the conserved histidine [69, 70]. HKs with such domain organisation comprise approximately 5% of all HKs [4]. Despite different domain arrangement and phosphotransfer mechanism, the CheA domains involved in autophosphorylation are homologous to those found in canonical systems and phosphorelays, which again emphasises the modular nature of TCSs.

1.2.1.2 HWE histidine kinase core

High level of sequence conservation across the CA domains facilitates identification of HK proteins in prokaryotic genomes. However, another form of HK core exists, which contains significantly different amino acid motifs within the CA domain. The HWE HK core was discovered in a search for bacteriophytochrome photoreceptor proteins and

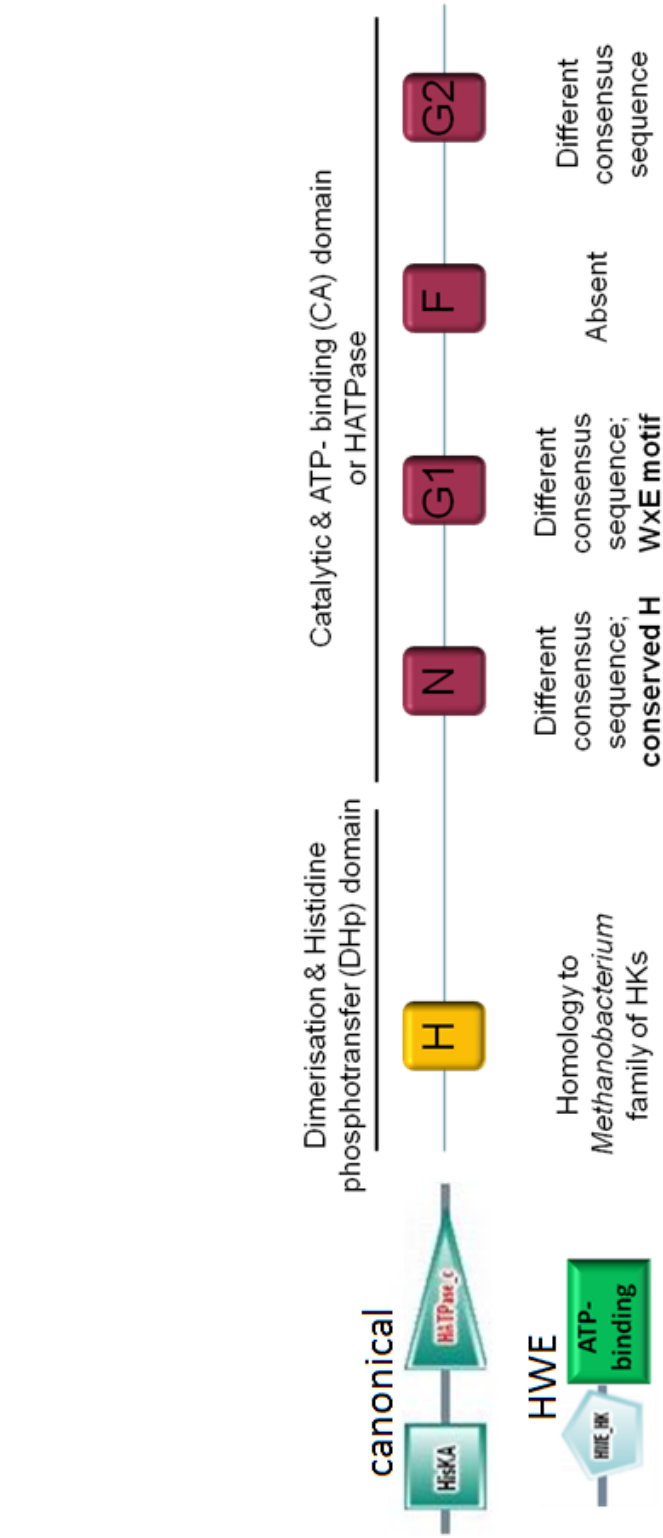


Figure 1.5: Comparison of the conserved sequence motifs in canonical and HWE HK cores. The boxes containing conserved residues are shown in order that they appear along HK core sequence. HK cores are represented using SMART graphical output [64].

shown to possess histidine kinase activity [71]. In HWE HKs the domain equivalent to that of DHp in classical HKs is distantly related to one uncharacterised class of HKs but the C-terminal portion is not annotated when scanning against SMART or Pfam databases. There are very different consensus sequences in place of the highly conserved N/G1/F/G2-boxes defining the ATP-binding cleft in canonical HKs, although some of the residues normally involved in nucleotide binding (N-box asparagine and G1/G2-box glycines) are present [60]. A comparison between canonical and HWE cores is given in Figure 1.5. Lack of an evident F-box may suggest that the ATP lid is absent or possesses different properties than lids in canonical HKs. The name of this HK class derives from three invariant residues - a histidine within the N-box and a tryptophan-X-glutamate motif within the G1-box.

HWE cores have been identified only in some bacterial phyla (α- and γ-proteobacteria) and are particularly enriched in rhizobial species. They are often linked to photosensing TCSs (bacteriophytochrome systems). Only a few representatives have been tested for activity so far [60, 71].

1.2.2 Histidine phosphotransfer (Hpt) domain

As already described above, the phosphorylatable histidine residue of CheA is located within the N-terminal Hpt domain and not in the dimerisation domain directly preceding the CA domain. Hpt domains are either present in chemotaxis HKs or in phosphorelays, where they mediate between two receiver domains [3, 4]. They receive the phosphoryl group from the upstream receiver domain but are incapable of autokinase or phosphatase activity; an increased number of phosphotransfer steps may provide more regulatory checkpoints and some Hpt domains or intermediate RRs can be targeted by specific phosphatases [4].

Phosphorelay Hpt domains can constitute C-terminal extensions of hybrid HKs or exist as separate proteins. Hpt domain architecture distantly resembles the four-helix bundle formed by DHp domains of HKs, but in most cases all four helices belong to one polypeptide; such domain exhibits an α-up-down topology [70, 69, 4]. Despite the presence of the invariant histidine residue and several other crucial amino acids, sequence conservation of Hpt domains is relatively low, which makes their identification challenging.

1.2.3 Receiver domains

RR receiver domains display high sequence similarity and share a common $(\beta\alpha)_5$ fold, with a five-stranded parallel β-sheet flanked by helices (Fig. 1.6A) [4, 43]. The active site aspartate is located at the start of a loop that follows β3 and most of the residues involved in phosphotransfer catalysis are positioned within the loop regions surrounding the active site (Fig. 1.6B). There are two acidic (D/E) amino acids close to the N terminus whose role is to coordinate the magnesium ion, and a conserved lysine residue following β5 which, in the active form of the receiver domain, is linked to the phosphate group via a salt bridge. In addition to several more-or-less strongly conserved residues, there are two amino acid positions that in all the well-characterised RRs seem to mediate the change from an inactive to an active state: Thr/Ser (β4) and Tyr/Phe (β5). Upon phosphotransfer, the Thr/Ser residue reorients and forms a hydrogen bond with the phosphate oxygen, while Tyr/Phe, normally exposed to the solvent or interacting with an associated effector domain, takes over the space vacated by Thr/Ser. This conserved mechanism is called Y-T coupling after the two most common residues found in the respective positions. Changes in the rotameric states of these two residues, along with varying extents of backbone displacements, relay the phosphorylation signal towards the α4/β5/α5 interface which mediates protein-protein or interdomain interactions.

As discussed in Section 1.1.3.1, activation of some RRs involves changes in the oligomeric state. Given the high level of conservation of receiver domains, it seems likely that heterodimers could form *in vivo*. However, *in vitro* studies showed that, similarly to HKs, RRs possess specificity determinants that direct oligomerisation [73]. It is unknown whether this dimerisation specificity is conferred by the receiver domain alone as the RRs tested contained associated effector domains. Interestingly, some heterodimers were observed, and this suggests that RRs may potentially be involved in cross-regulation at the level of dimer formation [73, 74].

In several cases of RRs with associated output domains the contacts between the receiver and output domains have been shown to have an inhibitory effect on autophosphorylation of the receiver domain with small-molecule phosphate donors [75]. It appears that extensive interdomain interactions lock the RRs in an autophosphorylation-incompetent state.

1.3 Origin, evolution and prevalence of TCSs

TCSs evolved from simple cytoplasmic one-component systems in response to the need for extracytoplasmic sensing [76, 16]. They are prevalent but not evenly distributed

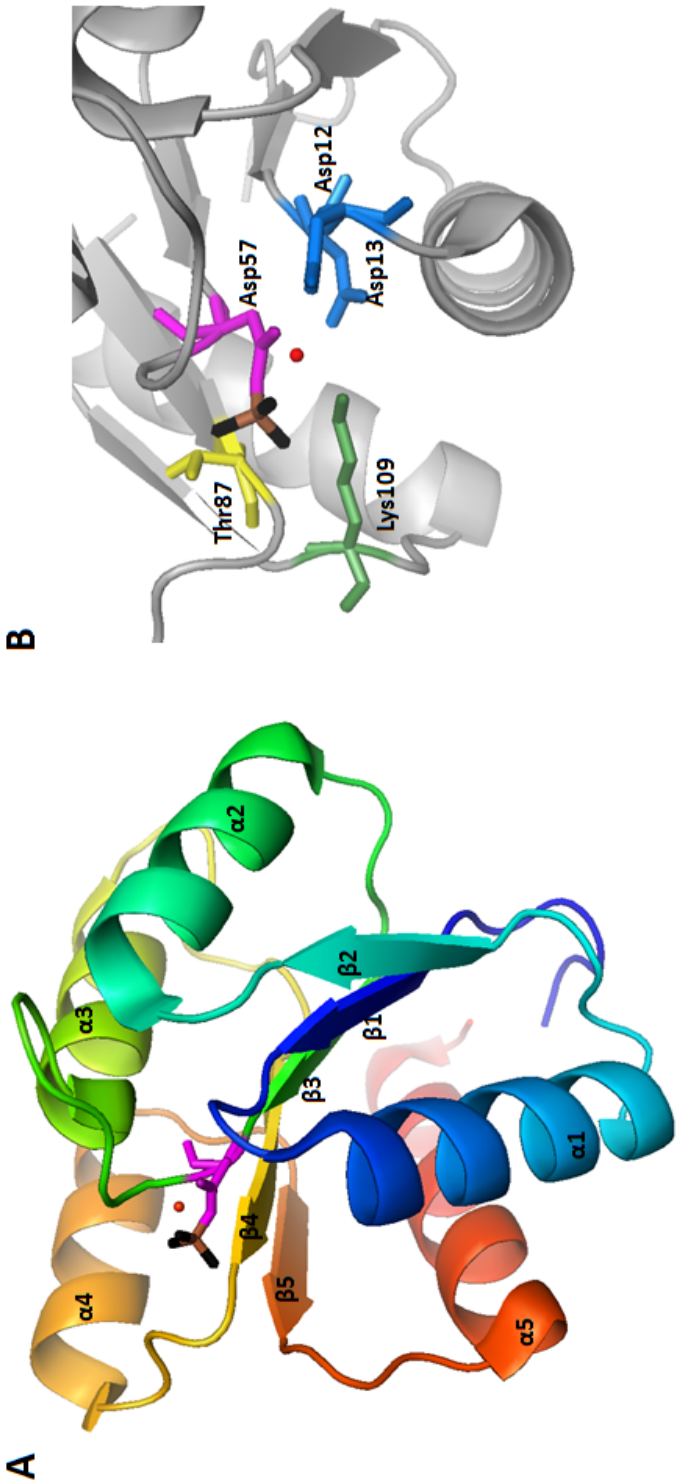


Figure 1.6: The structure of activated receiver domain and the active site. A - overall structure of CheY with bound phosphate mimic, beryllofluoride (brown and black sticks), and magnesium ion (red sphere), showing the central 5-stranded parallel β-sheet and two layers of helices. Phosphorylatable Asp57 is in magenta and the protein is presented using the rainbow colouring scheme. B - close-up of the active site showing Asp57 and four other highly conserved residues constituting the active site. PDB ID 1FQW. [72]

among microorganisms and are rare in eukaryotes [76, 16, 18]. Evolution of this type of signalling involved various genetic processes which allowed both TCS expansion and the creation of novel signalling proteins and pathways [77, 78]. HKs and RRs constitute large paralogous gene families and the specificity of HK-RR interaction is governed primarily by molecular recognition to ensure insulation of each separate pathway [17, 77].

1.3.1 Prevalence of TCSs and "bacterial IQ"

Distribution of TCSs has been frequently used as an argument for bacterial origin of this form of signal transduction [16, 18, 79, 80] - HKs and RRs are widespread in bacteria, except for certain species with reduced gene content, but they are not as well represented in archaea and eukaryotes. Only about half of the sequenced archaeal genomes and less than 30% of eukaryotic genomes encode TCS proteins [16], and HKs and RRs are virtually absent from the animal and protist kingdoms [18]. TCSs seem to have spread from bacterial species via horizontal gene transfer, which is evidenced not only by their prevalence in bacteria but also by the diversity they exhibit; eukaryotic and archaeal TCSs tend to be less diverse and one species often contains HKs or RRs belonging to the same type [80, 79]. Higher plants like *Arabidopsis thaliana* and lower eukaryotic species frequently possess several TCSs that, in conjunction with other signalling cascades, control various developmental pathways, such as ethylene response in plants or initiation of MAP kinase cascades in yeast [3, 18].

The number of both one- and two-component systems has been shown to increase with genome size of the microorganism [76, 81] but it is also affected by the lifestyle, metabolic complexity and the environment [81, 82, 83, 79]. The ability to sense intra- or extracellular changes, which includes the proportion of TCS genes, can be referred to as "bacterial IQ" [81]. Microbes utilising diverse energy sources or undergoing complicated developmental programmes usually contain numerous signalling pathways and therefore have a high IQ. There are both "introverts", primarily sensing changes occurring within the cytoplasm, and "extroverts", which monitor extracellular parameters. At the other extreme, prokaryotic species occupying stable environments, such as pathogenic or endosymbiotic bacteria, often possess minimal genomes that are not substantially regulated by the external stimuli and hence encode few or no TCSs [16, 81].

1.3.2 TCS gene expansion and diversification

There are two modes through which TCSs spread within and between species [82, 77]. Lateral gene transfer involves transmission of gene(s) between two organisms and is more likely to preserve the original functionality of a given protein or entire pathway. It constitutes a prominent mode of TCS acquisition for microorganisms and some species, like *E. coli*, rely primarily on this mechanism [82]. Vertical gene transfer, or lineage-specific expansion, occurs within the species and involves gene duplication, often followed by domain shuffling or gradual sequence differentiation from the parental gene. This expansion mechanism is frequently utilised by bacteria occupying complex environmental niches or possessing extensive metabolic networks [82]; it often introduces new domain arrangements and functions.

Both expansion modes are widespread, but there are several cases of species-specific preference for either horizontal or vertical gene transfer. In addition, gene fusions and fissions, which can accompany these two types of gene expansion, seem to be a relatively rare but important factor contributing to diversification of TCS proteins [78, 84]. The emergence of hybrid HKs and RRs most likely allows more complex regulation of the response by introducing new phosphotransfer checkpoints and frequently bridging TCSs with other signalling systems [78, 14, 3].

1.3.3 Hybrid RRs

Hybrid RRs, which are common in archaea and can be found in some bacterial species [16], are a particularly interesting type of gene fusion, as the activity of their C-terminal HK cores may be regulated by the N-terminal receiver domains. 25 proteins exhibiting this type of arrangement were identified during a search for proteins that contain the signalling helix, which is one of the structural linkers commonly found between the sensory region and the HK core of canonical HKs [28] (see Section 1.1.1.2). As the signalling helix can be associated with diverse signalling proteins and most likely serves to relay the signal arising from stimulus perception, it is conceivable that in such hybrid RRs the autokinase activity is regulated by the receiver domain in a phosphorylation-dependent manner.

1.3.4 Genomic organisation - operonic TCSs and orphan proteins

A common feature of bacterial genomes is that genes encoding proteins of similar function are frequently clustered together; this way they are more likely to be transferred as a group and produce a particular phenotype [85]. TCSs which are found in operons appear to have spread across different species as an intact pathway [77]. Gene co-localisation greatly facilitates identification of interacting pairs or groups of TCS proteins, but there are cases where no putative signalling partner is encoded in the vicinity of a TCS protein. Gene duplication, as described in the section before, is a frequent mode of TCS expansion and may involve an entire pathway, individual proteins or partial sequences. When only one of the two genes encoding TCS proteins is duplicated, it may undergo gradual differentiation but the orphan protein frequently retains phosphotransfer specificity with the cognate protein of the parental paralogue. If a process such as domain shuffling occurs, the orphan can acquire new sensory or regulatory capabilities and may become involved in cross-regulation. This is the likely explanation for the presence of several orphan Kin HKs in *Bacillus subtilis* which can all phosphorylate the same RR, Spo0F; converging multiple inputs or outputs allows fine-tuning of the adaptive response. Orphan proteins can account for a large proportion of the TCS complement and may possess atypical functions [86].

1.3.5 Phosphotransfer specificity

HKs and RRs constitute two of the largest paralogous gene families, with many bacterial cells possessing numerous TCSs working in parallel [3, 17, 77]. The significant level of conservation of the two basic signalling modules - HK core and receiver domain - raises the question of how only cognate pairs of proteins form productive interactions. Each pathway needs to be insulated in order to prevent cross-talk which would lead to cellular chaos.

There are cases where spatial or temporal separation of different systems helps maintain interaction specificity [17]. Some TCS proteins are only expressed at certain developmental stages or are confined to different parts of the cell. Scaffold proteins can also be involved in forming large protein complexes which keep pathway members away from the rest of the cell, as observed in the case of chemotaxis systems. About a quarter of all HKs are hybrids [4] and the spatial constraint of the receiver domain with respect to the HK core may reduce the chances for cross-talk. In fact, the receiver domain of a hybrid HK purified as a stand-alone domain protein exhibited phosphotransfer

promiscuity, which suggests that physical constraints significantly decrease the selective pressure [87]. Finally, the considerably higher concentration of RRs with respect to HKs observed *in vivo* may be an evolutionary strategy to minimise potentially detrimental interactions [17]. However, all the measures described do not seem sufficient to ensure specific protein interactions in most cases [17].

Even though none of the available HK-RR complex structures capture the phosphotransfer reaction [32, 37], formation of the phosphotransfer interface has been shown to rely on molecular recognition [88, 89, 67, 90]. HKs exhibit strong kinetic preference for their cognate RRs which can be tested by means of phosphotransfer profiling [89] - high levels of phosphotransfer between cognate proteins occur within a very short period of time (seconds or minutes), while phosphorylation of RRs by non-cognate HKs requires significantly longer incubation times. TCS proteins have coevolved so that each mutation within one protein that disrupts the binding interface needs to be compensated by a mutation on the other protein, or the initial mutation is removed by negative selection [17, 67]. Identification of the crucial interface residues allowed *in vitro* alterations of interaction specificity within some systems, which usually entailed a few site-directed mutations and partial sequence grafting within the base of the DHp domain [67]. Rewiring the specificity by introducing changes to a receiver domain sequence has also proved successful [91, 90]. However, specificity determinants in some TCSs may be more complex to predict based on the sequence alone [67, 91]. In the context of the evolution and diversification, there may be certain constraints on the mutational possibilities available to TCSs - substitutions of some of the specificity conferring residues may abolish phosphotransfer ability and/or allow cross-talk to closely related systems, and would therefore be counteracted by natural selection [91].

TCSs in different organisms are subject to different selective forces and may need to differentiate accordingly. It has been reported that certain mutations that have occurred in the course of bacterial evolution were also driven by the need to eliminate cross-talk [92]. Such adaptive mutations were necessary in cases where one organism possesses two closely related TCSs, while another species utilises only one of these systems and therefore does not experience any cross-talk. In intrinsically branched systems, different interaction specificities between HK-RR pairs may account for different kinetic phosphotransfer preferences, thus allowing complex regulation of various processes [93].

1.4 Bacteriophytochrome signalling systems

Photosynthetic organisms need a way of measuring light intensities in order to regulate photosynthetic apparatus synthesis as well as various photobiological processes [94]. In higher plants flowering times, seed germination, response to shade and several other vital processes are controlled by phytochromes: red and far-red light photoreceptors which interconvert between active and inactive forms depending on light intensity. Cyanobacteria and other photosynthetic microorganisms have been shown to utilise phytochromes for phototaxis [95], expression of light-harvesting and photosystem apparatus [96, 97], and protection against bright light [98].

These photoreceptors appear to have bacterial ancestry [99, 100] and some of them have evolved to detect different wavelength ranges within the red spectrum [97]. They act as binary switches between two light absorbing forms; the one absorbing red light (Pr) is the predominant ground state [94] but some phytochromes assume the form absorbing far-red light (Pfr) as the ground state [71]. Signal transmission evoked by phytochromes reflects the red and far-red light intensity ratio [101]. Plant phytochromes are thought to act as serine/threonine protein kinases [102, 101], although some evidence suggests that they interact allosterically with transcriptional factors and potentially other proteins [94]. Bacterio- and cyanobacterial phytochromes belong to the HK family of proteins and have been shown to induce two-component signal transduction [103, 99, 104].

Phytochromes are homodimeric proteins with an N-terminal photosensory domain and the transmitter domain located at the C-terminus [101]. Typical sensing regions comprise three domains - PAS, GAF (cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA) and PHY (phytochrome), which may have common ancestry as evidenced by topological similarities [105]. All phytochromes require a bilin chromophore which is covalently attached to the GAF domain [101, 98]. Plant phytochromes bind phytochromobilin via an invariant cysteine residue, while cyanobacterial photoreceptors also use cysteine for attachment but utilise phycocyanobilin as the chromophore [101].

1.4.1 Structure and function of bacteriophytochrome photoreceptors (BphPs)

BphPs differ from their plant counterparts in terms of the C-terminal transmitter domain, the type of chromophore associated with the sensor domain, and the way the chromophore is attached to the protein. They act as light-dependent HKs with a biliverdin chromophore [98] which is attached to a cysteine located close to the amino

terminus [106, 97]. Depending on the state (Pr or Pfr), kinase activity may be up- or down-regulated and differences between the levels of autokinase activity in Pr versus Pfr form seem to vary across species [71, 107, 108]. The three-dimensional structure of the entire photosensory region of a *Pseudomonas aeruginosa* BphP revealed a homodimeric structure with PAS and GAF domains forming the chromophore-binding pocket [105]. A model based on that structure postulated that the last helix of the PHY domain extends into the first helix of the HK core, thus allowing conformational changes within the photosensory region to be transmitted directly to the module containing the phosphorylatable histidine. Electron microscopy data from another study showed that BphPs form homodimers via extensive interfaces contributed by the HK core, as well as all three domains from the photosensory module [109]. It appears that light-dependent changes within the chromophore would affect the GAF/PHY and dimerisation interfaces, leading to changes in the orientation of the HK catalytic domain with respect to the active site histidine.

BphPs have been found in both photosynthetic and some heterotrophic bacteria [98]. Microbial genomes typically encode more than one BphP-linked signal transduction system; in *Rhodopseudomonas palustris* two BphPs monitor changes in two different wavelength ranges [97], while in *Agrobacterium tumefaciens* two BphPs possess contrasting photobiological properties (different ground states) and may thus exhibit opposing roles [71]. In addition, some BphPs have undergone diversification, as in the case of *R. palustris* BphP4 - in some strains this protein has lost the ability to bind the chromophore and instead evolved into a redox sensor [110]. One study also implied that along with red/far-red light detection, BphPs may sense temperature [108]. The output of BphP signalling cascades in non-photosynthetic bacteria still largely remains a mystery, but is likely to involve changes in gene expression or motility as in the case of several characterised BphPs from photosynthetic bacteria and cyanobacteria [110, 97, 96, 95] .

1.4.2 Bacteriophytochrome signalling cascades

There are only a few signal transduction systems involving BphPs that have been fully or partially characterised to date. Some BphPs do not possess any signalling regions at their carboxyl termini, suggesting protein-protein interactions as a mode of transmitting the signal [96], but in most cases they utilise two-component transmitter modules. The primary obstacle in identifying responses triggered by BphPs is the lack of associated effector domains - the cognate RRs tend to be stand-alone receiver domain proteins [98, 71, 97], similarly to cyanobacterial RRs associated with phytochromes [103].

In the photosynthetic bacterium *R. palustris* BphP4 has been shown to act as either

a photoreceptor or a redox sensor, depending on the clade, but in each case signalling involved an orthodox RR with a C-terminal DNA-binding domain. Single domain RRs, however, do not provide any information on the downstream signalling pathway. In *R. palustris* both BphP2 and BphP3 were able to phosphorylate one stand-alone receiver domain RR but neither phosphorylated the other single-domain protein or a hybrid RR, both of which are encoded next to the photoreceptor genes; the signalling pathways from the two BphPs therefore seem to converge [97]. The hybrid RR present in *R. palustris* features an unusual domain arrangement, with the receiver domain located N-terminal with respect to a HK core, suggesting that the autokinase activity may be controlled by phosphorylation of the receiver domain by BphPs. Such domain organisation has also been observed in *A. tumefaciens* ExsG protein [71] (see next section).

1.4.3 Bacteriophytochrome signalling in *Agrobacterium tumefaciens*

There are two BphPs in *A. tumefaciens* which, as previously mentioned, possibly regulate two different processes as their photobiological properties are contrasting [71]. Interestingly, there is evidence to suggest that one of them, BphP1, may monitor temperature as well as light changes [108]. BphP1 is encoded within a gene cluster that incorporates two stand-alone receiver domain proteins, ExsF and AgR, and ExsG, an atypical hybrid HK with N-terminal receiver domain [71]. The genes are unlikely to constitute an operon given their overlapping reading frames and the fact that *ExsF* points in the opposite direction. The other photoreceptor, BphP2, is encoded elsewhere in the genome and constitutes a hybrid sensor kinase, with a receiver domain appended at the carboxyl terminus. There are no obvious two-component protein genes in the vicinity of BphP2 gene. Figure 1.7 depicts the organisation of the two BphP gene clusters and domain arrangements of the proteins.

Both BphP2 and ExsG employ the uncommon HWE HK core and their autokinase activities have been demonstrated along with the activity of BphP1 which contains the canonical form of HK core [71]. In addition, out of the two photoreceptors only BphP1 was capable of phosphorylating AgR, while BphP2 specifically phosphorylates its appended receiver domain. The activities of ExsF and of ExsG receiver domain have not been tested.

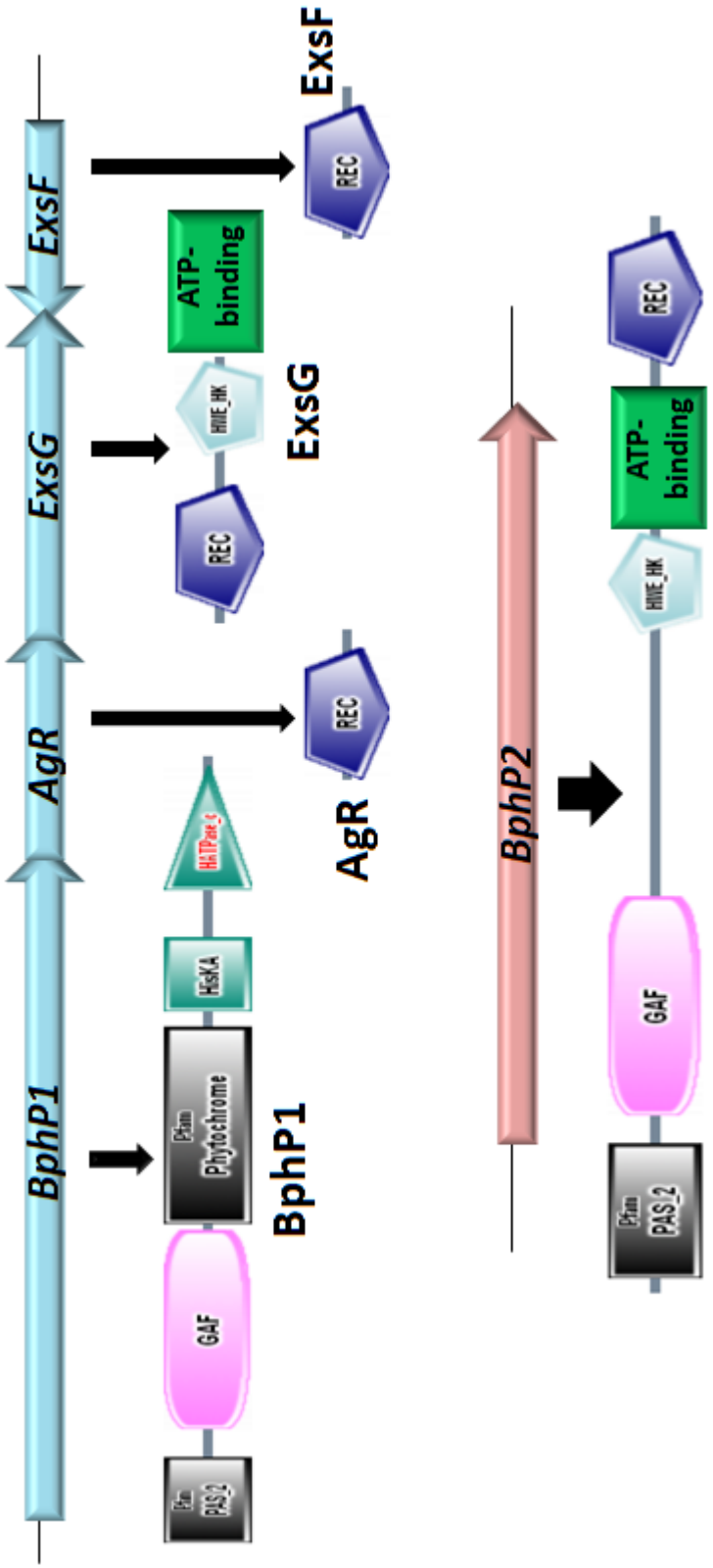


Figure 1.7: Gene and domain arrangement of *A. tumefaciens* BphP1 and BphP2 cluster proteins. Direction of genes is indicated by the arrowheads, genes are not to scale. Domain organisation of the proteins is represented using SMART graphical output [64]. The ATP-binding region, not detected by SMART in *A. tumefaciens* ExsG and BphP2, is shown as a green box.

1.5 *Rhizobium* NT-26 and the significance of its two-component "signalome"

Rhizobium NT-26 is a member of the *Agrobacteria/Rhizobia* genus but, in contrast to its typical soil relatives, it was isolated from a gold mine in Australia [111]. This biflagellate Gram-negative bacterium is capable of utilising arsenite as an electron donor and constitutes the most robust arsenite oxidiser characterised so far. In addition to chemolithoautotrophic arsenite oxidation, NT-26 is capable of sulfur oxidation and respiration with both nitrate and nitrite (J. Santini, personal communication). The expression of its arsenite oxidase is regulated by a canonical TCS, AioSR, but there is evidence to suggest that additional signalling pathways downstream of AioS might control resistance to arsenite [112]. NT-26 genome sequencing was finished during the course of this project but the genome sequence was not published before its end.

With the already demonstrated ability to manipulate the specificity of HK-RR interactions [67, 91, 90], and to create functional chimaeric proteins [6, 7, 9], it is becoming possible to design complex signalling circuits and recombinant microorganisms with unusual and programmable properties [9, 10]. There is a vast diversity of sensory domains associated with TCSs that have not yet been structurally and functionally characterised; some of these domains could be exploited in constructing novel signalling pathways or biosensors. Characterisation of proteins with uncommon domain arrangements and non-canonical domains involved in phosphotransfer may be essential for designing more complex signal transduction circuits. However, to take full advantage of the intrinsic modularity and plasticity of TCSs, it is necessary to expand the pool of available signalling modules. Therefore, attempts should be made to identify and characterise various TCSs, particularly those employing unusual domains and domain arrangements, in terms of function and signalling mechanism. Furthermore, the mechanisms of interdomain signal propagation as well as the determinants of phosphotransfer specificity need to be fully elucidated to allow effective coupling of different modules. Finally, in the context of whole-cell engineering, the possibility of cross-talk and cross-regulation between different TCSs would also need to be explored.

From the point of view of synthetic biology, organisms like NT-26 constitute potential sources of novel signalling units. Given that this microbe possesses several metabolic strategies and has adapted to an environmental niche rich in metals and toxic compounds, it seemed likely to encode unsual sensor domains, proteins and signal transduction systems. In the context of two-component signalling, the complexity of the environment would be reflected primarily in the number and diversity of transmem-

brane and periplasmic-sensing HKs. The set of signalling proteins encoded by NT-26 genome - the "signalome" - would therefore be expected to differ from, and potentially to be greater than, the "signalomes" of closely related species, of which most are plant pathogens or symbionts.

1.6 Thesis aims

The starting goal of this project was to identify and characterise all two-component signalling proteins encoded within NT-26 genome and identify any unusual or otherwise interesting systems and proteins. Chapter 3 describes the analysis of NT-26 TCSs and comparison of the entire NT-26 "signalome" to "signalomes" of its close relatives. It also summarises the results of small-scale expression trials of members of selected putative TCSs. One of these systems was to be further characterised provided that all its members could be purified as soluble proteins.

During the analysis of NT-26 two-component "signalome" the homologues of *A. tumefaciens* BphP1 and BphP2 were discovered. Identification of photorceptors in a bacterium occupying a dark environment was intriguing, and the likely complexity of the putative signalling cascade initiated by BphP1 made this TCS a good candidate for further characterisation. In addition, the responses evoked by BphPs have not yet been elucidated. Therefore the second aim of this project was to characterise the signalling pathway downstream of BphP1 and to potentially identify the response(s) it triggers. Chapter 4 describes purification of the putative members of the pathway, determination of signalling direction and attempts to establish the final output.

The domain arrangement of one member of BphP1-mediated pathway, ExsG, raised questions about the signalling role of this protein - given the presence of an N-terminal receiver domain and C-terminal HK core, it was not obvious whether ExsG acts as a RR or a HK. In addition, the HWE catalytic core present in ExsG has only been subjected to basic biochemical tests [71, 60], but at the time when this thesis was written no structural or biochemical data were available on this type of HK core. The third aim of this project was therefore to characterise the native state of ExsG (Chapter 5) and elucidate its role in signal transmission (Chapter 6).

Chapter 7 provides conclusions and discussion on the biological significance of NT-26 two-component "signalome" and the signalling pathway downstream of BphP1. Suggestions for future studies and potential applications of ExsG and its associated TCS are also given therein.

Chapter 2

Materials and methods

This chapter lists materials and describes techniques used in this project. Unless stated otherwise, the chemical reagents were purchased from VWR, Sigma-Aldrich or Fisher Bioreagents. All bacterial cells used in cloning and protein expression were *Escherichia coli* strains.

2.1 Bioinformatic analyses

2.1.1 Identification and annotation of TCS genes in *Rhizobium* NT-26 genome

The sequencing of NT-26 genome, conducted by Genoscope - Centre National de Sequencage (French National Centre for Sequencing), was accomplished just before the submission of this thesis (GenBank accession code: FO082820.1). Access to the genomic data throughout the project was provided by Genoscope[1]. The genome website enabled searches using associated keywords and/or BLAST [113]. Searching for TCS proteins was performed iteratively, with newly identified proteins or their domains serving as queries for subsequent searches until no new proteins with receiver domains or kinase cores were identified.

Initial rounds of searches for TCS proteins involved scanning the genomic sequence with NT-26 AioS HK core, AioR receiver domain, and with the full sequence of each of these proteins. The phosphotransfer domains of identified proteins were then used as queries for subsequent rounds of BLASTp searches against NT-26 genome. The sequences of domains commonly found in HKs (GAF, PAS and HAMP) and RRs (Trans_reg_c) were also used as queries. In addition, several full and partial *E. coli* protein sequences were scanned against NT-26 genome: CheY (UniProt [114] accession

[1] https://www.genoscope.cns.fr/spip/

number P0AE67), CheA (P07363), EnvZ (P0AEJ4), OmpR (P0AA16) and Hpt domain portion of CheA. The keyword search tool was also used to detect TCS-related gene annotations.

The presence of phosphotransfer domains in identified protein sequences was verified by submitting them to the SMART webserver[2] (simple modular architecture research tool), which allows simultaneous scanning against SMART [64] and Pfam [65] databases. Multiple sequence alignment created using default parameters with MUSCLE online tool[3] (European Bioinformatics Institute [115]) was examined to confirm the presence of the phosphorylatable residues within the predicted HK cores and receiver domains. Pairs or groups of proteins encoded in operons or gene clusters were assigned into putative signalling systems. The sequences of proteins belonging to the assigned systems were used as queries in BLAST and PSI-BLAST (National Centre for Biotechnology Information [113]) searches against the nr (non-redundant protein sequence database) in order to identify homologues. Function of the system was assigned if all system members matched, in terms of at least 50% overall sequence identity and domain architecture, homologous proteins in other organisms comprising a pathway whose function has been experimentally verified. In addition, the sensory portions of NT-26 HKs also had to share at least 50% sequence identity with the respective regions of homologous proteins. In the case of pathways potentially involving more than two proteins, the operonic or gene cluster arrangement of NT-26 proteins also had to correlate with that of homologous proteins.

2.1.2 Multiple sequence alignments

All alignments were created using MUSCLE [115] and viewed in Jalview version 2 [116] environment with default ClustalX colouring scheme applied. Under this scheme, explained in Appendix B, Table B.1, the residues are coloured according to the amino acid profile at each position. For example, lysine and arginine appear red when at least 60% of the residues at a specified position of the alignment are Lys/Arg. Pairwise percentage sequence identity were calculated using Jalview 2.

2.1.3 Structural predictions

Coiled-coil regions were predicted using the programme COILS[4] [117] with default parameters; the results obtained from scanning with a 21 residue window were used. Sec-

[2]http://smart.embl-heidelberg.de/
[3]http://www.ebi.ac.uk/Tools/msa/muscle
[4]http://embnet.vital-it.ch/software/COILS_form.html

ondary structure predictions were calculated by the programme PSIPRED[5] [118] and, where specified, by Jpred3[6] [119]. The pGenTHREADER fold recognition tool [120] was accessed through PSIPRED webserver interface. Identification of structural homologues was performed in two ways: (i) the sequences were used as queries in BLASTp searches against the Protein Data Bank (PDB), or (ii) were scanned against the CATH (class, architecture, topology, homology) domain superfamily Hidden Markov Model (HMM) profiles using Gene3D[7] (CATH Domain Homologue Recognition Web Service) [121]. The threaded structure of ExsG receiver domain (residues 6 - 130) was generated by the automated homology modelling server BioSerf [118].

2.1.4 Figures

All figures showing protein structures as well as structure superimposition were generated using the PyMOL Molecular Graphics System [122].

2.2 DNA manipulation, purification and extraction

2.2.1 Agarose gels

All gels contained 1% agarose (w/v) (Melford) and 1 x TAE buffer (40 mM Tris-Cl pH 8.0, 20 mM acetic acid and 1 mM EDTA). 3 μl of ethidium bromide were added per 50 ml of melted agarose solution before casting a gel. Gels were run at 100 mV for 40 min in 1 x TAE buffer and photographed under UV light. The molecular weight marker used was Hyperladder I (Bioline) and the samples were mixed with 5x DNA loading dye (Bioline).

2.2.2 DNA concentration measurements

The concentration of DNA was determined by measuring the absorbance at 260 nm using NanoDrop spectrophotometer (Thermo Scientific). It was calculated based on the Beer-Lambert law:

$$A_{260} = \varepsilon l c$$

[5] http://bioinf.cs.ucl.ac.uk/psipred/
[6] http://www.compbio.dundee.ac.uk/www-jpred/
[7] http://gene3d.biochem.ucl.ac.uk/Gene3DScanSvc/SuperfamilyScan

where A_{260}- nucleic acid absorbance at 260 nm, ε- molar extinction coefficient at 260 nm $[\frac{M}{cm}]$, c - nucleic acid concentration $[M]$, l - path length $[cm]$.

2.2.3 Lysogeny Broth (LB) media and agar plates

LB, containing 10 g/L tryptone, 5 g/L NaCl and 5 g/L yeast extract (LB broth, Sigma-Aldrich) was the growth medium for liquid cultures of *E. coli* strains, autoclaved and supplemented with appropriate antibiotic where necessary. LB agar Lennox capsules (MP Biomedicals) were used to make LB agar which in addition to tryptone, NaCl and yeast extract contains 15 g/L agar. After autoclaving, LB agar was poured neatly into Petri dishes. Kanamycin or ampicillin (Melford), as appropriate, was added to make the final concentration of 50 μM (kanamycin) or 100 μM (ampicillin) in the total volume of both LB media or agar.

2.2.4 Plasmid extraction and purification

Plasmids were extracted using QIAprep Miniprep Kit (Qiagen), which is based on the alkaline lysis method. Pelleted cells are lysed under alkaline conditions, and the solution is neutralised in order to provide high-salt binding conditions. DNA is bound by special silica membrane of the QIAprep spin columns during centrifugation, then washed to remove salts and the plasmid is eluted from the membrane using water.

2.2.5 Genomic DNA extraction by alkaline lysis

A scoop of *Rhizobium* NT-26 cell pellet was homogenised in 3 ml of solution A containing 25 mM Tris-Cl pH 8.0, 10 mM EDTA, 50 mM glucose, with 5 mg lysozyme (Sigma) and 20 mg ribonuclease A (Qiagen) added fresh to the solution, and incubated for 5 min at room temperature. 1.5 ml of solution B (0.02 M NaOH, 1% w/v SDS) was then added, the contents mixed by inverting and incubated for further 5 min at RT. After adding 1.2 ml of solution C (calcium acetate pH 4.8) the tube was incubated for 15 min on ice and centrifuged for 30 min at 17000 x g. The supernatant was transferred to a fresh tube and centrifuged again. 0.6 ml of ice-cold isopropanol was added to the supernatant transferred to a fresh tube, and after gentle inverting threads of DNA began to appear in the solution. The tube was incubated overnight at 4° C and centrifuged at 17000 x g for 30 min. The supernatant was removed and 1.5 ml of ethanol was added to the

pellet. The tube was inverted gently several times, centrifuged at 17000 x g for 30 min, and the supernatant was again removed. The pellet was allowed to air dry thoroughly and finally 30 μl of buffer TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA) was added and the pellet was gently resuspended. Genomic DNA was stored at 4^o C.

2.2.6 DNA purification from agarose gels

DNA fragments of appropriate size were excised from agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen). The gel slice was melted in the binding buffer, applied to a spin column and centrifuged. DNA was washed with ethanol-containing wash buffer and eluted in water.

2.2.7 DNA purification from enzymatic reactions

Following enzymatic digestion the DNA was purified using QIAquick PCR Purification Kit (Qiagen). The reaction was mixed with the binding buffer, DNA was bound to the membrane, washed and eluted in the same way as in gel extraction.

2.2.8 Restriction enzyme digestion

NcoI, BspHI, EcoRI, NdeI and HindIII (New England Biolabs) restriction enzymes are all compatible in double digestion reactions at 37°C. The reaction buffer was also supplied by New England Biolabs. For each 20 - 30 μl reaction approximately 20 units (1 - 2 μl) of the appropriate enzymes were used. Vector digestions were incubated for 2 hours, while insert digestions for at least 5 hours or overnight to ensure maximum efficiency. Diagnostic reaction volumes were 10 - 15 μl, with 5 - 10 units of the appropriate enzyme and incubation times of 1 - 2 hours. NdeI enzyme was used for cutting inserts out of the vector where ligating BspHI-cut insert into NcoI-cut vector disrupted the NcoI site.

2.2.9 Preparation of the linearised vector DNA

pET30a plasmid (5.5 kbp) was used for cloning all genes and transformations of both DH5α and BL21 DE3 pLysS cells. The vector has a kanamycin resistance gene, T7 promoter system and two His-tags, N- and C-terminal of the multiple cloning site (see Appendix A, Fig. A.1). Glycerol stock of the vector-carrying *E. coli* DH5α strain

was used to inoculate 5 ml of LB with 50 μg/ml of kanamycin, which was incubated overnight at 37° with shaking. The plasmid was extracted from 3 - 5 ml of cells using QIAprep MiniPrep Kit and digested with restriction endonucleases NcoI and HindIII. The digestion reaction was run on a 1% agarose gel and linearised plasmid DNA was excised from the gel and extracted using QIAquick Gel Extraction Kit.

2.2.10 Primer design for gene amplification

The primers were designed to either amplify the entire gene sequence or a part of it that encodes the phosphotransfer domain (receiver domain or HK core). The latter approach was necessary in the case of HKs which contained transmembrane regions. It was also applied to RRs with AAA+ ATPase domains (DctD and NtrC) and BphP1. Care was taken to ensure that the amplified gene portion would encode intact domains and the primers were designed such that the truncated protein construct would contain several residues of the interdomain linker.

CONSTRUCT	FORWARD PRIMER (5'-3')	REVERSE PRIMER (5'-3')
ExsG	gc**tcATGa**ctGTGCATAATCAGCTTCGC (67)	cg**aagctt**CTATCCGGAAAGATCTGCG (67)
ExsG-HK (130-336)	gc**tcatga**ggGAACGGACAGAGCGTC (70)	cgg**aagctt**aTCCGGAAAGATCTGCGAGG (70)
ExsG-REC (1-129)	gc**tcATGa**ctGTGCATAATCAGCTTCGCC (68)	cg**aagctt**aTGCCTCCGCGAGCGCC (71)
BphP1-HK (462-703)	gc**ccatgG**TGCTCCGCCCCTC (70)	gcg**aagctt**TCAGGCAATCGCTCTTCCC (70)
ExsF	gg**ccATGg**ctAATGTTCTGGTTCTGGAAG (65)	gg**aagctt**TCAGTTGCTGGCGGCAC (66)
AgR	gg**ccatgG**TGCCTGAACTGAGACC (66)	gg**aagcTT**ATGCACGTCCATTGCGCAG (66)
Hpt	gag**ccATGG**CCGCAACAGCAATTG (66)	gcgg**gaattc**TTATCGGCAAAGCTTGAG (67)

Table 2.1: Primers used for amplification BphP I system genes (partial and full-length sequences) and Hpt domain protein. Restriction enzyme sites (AAGCTT- HindIII, TCATGA - BspHI, CCATGG - NcoI, GAATTC - EcoRI) are shown in bold. Bases encoding protein sequences are in capital letters. The stop codon incorporated within ExsG-REC gene sequence is underlined. The numbers in brackets denote the average melting temperature in °C.

Each forward and reverse primer comprised a stretch of bases at the start or end of a gene (or region, in the case of truncated genes) and 5' flanking regions. Forward

primers contained NcoI (CCATGG) or BspHI (TCATGA) recognition sites; BspHI site was introduced if an NcoI site was present within the sequence to be cloned. Reverse primers had a HindIII (AAGCTT) recognition site at their 5' end. An exception was Hpt domain protein primer which incorporated an EcoRI cleavage site (GAATTC) as NT-26 Hpt sequence contains a HindIII site; the plasmid was digested accordingly. Two to three bases were added before each restriction enzyme recognition site. The portion of the coding sequence incorporated in the primer was selected in a way to ensure that the last base of the primer was a C or G. Stop codons were incorporated before the coding sequence of the reverse primers for the purpose of generating the C-terminal truncations (ExsG-REC, DctD and NtrC). The primers were matched in terms of melting temperature where possible. The portions of gene sequences contained within primer sequences were used as queries against NT-26 genome to ensure their length was sufficient for specific amplification. All primers were synthesised by MWG Eurofins. The primers used for cloning BphP1 cluster genes and their portions, as well as Hpt domain protein, are listed in Table 2.1; all other primers are listed in the Appendix, Table A.1.

2.2.11 Polymerase chain reaction (PCR)

PCR reactions,except for site-directed mutageneses, were performed using Phusion™ High Fidelity DNA polymerase (Phusion™ High Fidelity PCR Kit, Thermo Scientific) with 10-12% DMSO (v/v) in each reaction. 50 μl reactions contained 1 μl of the enzyme, 1 μl of NT-26 genomic DNA, 10 μl of Phusion 5x buffer, 2 μl of each 100 pmol/μl primer stock, 5 μl of dNTPs (2.5 mM stock), 5 - 5.5 μl DMSO and water. Thermal cycling was performed in Techne TC-312 thermocycler and involved 5 min initial denaturation at 95° C, then 35 cycles of (i) 30 sec denaturation, (ii) 30 sec primer annealing at 59 - 65°, (iii) 30 - 60 sec primer extension at 72°, followed by 10 - 20 min final extension. Extension time depended on the length of amplicon, while the temperature for primer annealing step depended on primer melting temperatures.

2.2.12 Insert preparation and ligation into the vector

All amplified DNA sequences were run on 1% agarose gels (with 3 μl of ethidium bromide per 50 ml of gel); bands of corresponding sizes were excised and extracted using QIAquick Gel Extraction Kit with elution volume of 30 μl. The DNA was digested with NcoI/BspHI and HindIII in a 30 μl reaction. The buffer and enzymes were removed using QIAquick PCR purification kit with elution volume of 30 μl .

The 20 μl ligation reactions contained 1 μl (1000 units) of T4 DNA ligase (New England Biolabs) and 2 μl of ligase buffer. Depending on the concentration of NT-26 DNA to be ligated into the vector, different ratios of insert : vector amounts were necessary. Usually, 1.5 - 3 μl of 60 ng/μl vector were used and the ligation mix was topped up to 20 μl with 20 - 50 ng/μl insert. The reactions were incubated at room temperature overnight.

2.2.13 Transformation of *E. coli* DH5α cells and screening

Subcloning Efficiency^{TM} DH5α competent cells (Invitrogen) were used according to the protocol provided. Each transformation required 50 μl of competent cells, thawed on ice, to which 1 - 4 μl of the ligation reaction were added. Cells were incubated on ice for 30 min, heat-shocked at 42^{o} C for 30 seconds and kept on ice for 2 min. Pre-warmed LB media (950 μl) were subsequently added and the tubes were incubated for 1 hour at 37^{o} C with shaking. Following incubation, 80 and 180 μl of cells were spread onto pre-warmed agar plates containing 50 μg/ml of kanamycin. The plates were incubated at 37^{o} C overnight.

The colonies were tested for the presence of insert by colony PCR. In 12.5 μl volume PCR reactions contained the appropriate primers, Phusion^{TM} High Fidelity DNA polymerase and its associated buffer, dNTPs, DMSO and water in the same proportions as in 50 μl reactions. For all NcoI-digested inserts the primers used for PCR amplification were used; T7 and T7 terminator primers were used for sequences in which NcoI RE site has been lost due to BspHI insert digestion. A colony was picked with a pipette tip, dipped several times in the PCR mix and placed into a tube with 5 ml of antibiotic-containing LB media. 2 - 5 colonies were picked from each transformation. PCR reactions were performed according to the standard protocol (see Section 2.2.11) and the products were run on 1% agarose gel. If a band of appropriate size was visible on the gel, the corresponding tubes with inoculated media were incubated overnight at 37^{o} C with shaking and plasmid DNA was extracted the next day using QIAprep Miniprep Kit; isolated plasmid DNA was digested for an hour and separated on 1% agarose gels. Due to false positive/negative anomalies encountered in some cases, at least two colonies from each transformation reaction were used to inoculate overnight cultures and were examined by double digestion. The list of all plasmids generated in this project is given in Table 2.2.

Plasmids shown to contain inserts of appropriate sizes were sent for sequencing, using T7 promoter and T7 terminator primers, to MWG Eurofins. The returned se-

quences were used as queries for BLAST searches against NT-26 genome. In the case of mismatch, the chromatograms were manually examined and sequences further verified.

2.2.14 Transformation of *E. coli* BL21 (DE3) cells

Plasmids shown to contain correct gene sequences were transformed into the expression strain BL21 (DE3) pLysS (Agilent Technologies) using the same protocol as for DH5α transformation (Section 2.2.13); 50 μl of transformed and grown cells were plated on kanamycin plates (pLysS carries chloramphenicol resistance gene but it was not selected for). The plates were incubated overnight at 37°C. The next day 5 ml LB with antibiotic was inoculated using one of the colonies and incubated overnight.

2.2.15 Transformation of *E. coli* ArcticExpress^{TM}(DE3) cells

Expression of NT-26 Hpt domain was attempted using ArcticExpress^{TM} (DE3) competent cells (Agilent Technologies). These cells contain chaperonins from a psychrophilic bacterium which are active even at reduced temperatures and assist with protein folding. Transformation was performed as for the other strains with the exception that β-ME (provided with the cells) was added prior to the addition of the plasmid and the heat-shock was 20 seconds long as recommended by the manufacturer.

2.2.16 Site-directed mutagenesis

Five specific single amino acid substitutions were introduced into ExsG full-length sequence. Primer design and mutagenesis were performed according to the protocol provided with the QuikChange Lightning^{TM} Site-Directed Mutagenesis Kit (Agilent Technologies) that was used for all mutagenesis reactions. PCR reactions were performed in Techne TC-312 thermocycler. Mutagenic primers were synthesised by MWG Eurofins and their sequences are given in Table 2.3.

Each 50 µl PCR reaction mix contained 1 µl QuikChange Lightning^{TM} Enzyme, 1x reaction buffer, 125 ng of forward/reverse primer, 1 µl of dNTP mix provided, 1.5 1 µl QuikSolution reagent and approximately 50 ng of the plasmid incorporating ExsG gene. After a 2 min denaturation step at 95° the plasmid DNA was subjected to 18 cycles of denaturation (20 seconds at 95°C), primer annealing (10 seconds at 60°C) and primer extension (3 min at 68°C) steps followed by final extension for 5 min at 68°C. To each amplification product 2 µl of the provided DpnI enzyme were added and the mixtures were thoroughly mixed and incubated for 5 min at 37°C to digest the parental DNA.

Protein ID (name)	Plasmid stock name	Portion cloned	RE sites for excision	Insert size (bp)
152	FixL	1-639 (full-length)	NcoI, HindIII	1920
153	FixJ	1-244 (full-length)	NcoI, HindIII	740
$917_{241\text{-}468}$ (FeuQ)	FeuQ	241-468 (HK core)	NcoI, HindIII	690
916 (FeuP)	FeuP	1-223 (full-length)	NcoI, HindIII	675
$925_{186\text{-}470}$	RagB	186-470 (HK core)	NdeI, Hind III	1000
924	RagA	1-251 (full-length)	NdeI, Hind III	900
1631 (NtrB)	NtrB	1-382 (full-length)	NdeI, HindIII	1300
$1632_{1\text{-}157}$ (NtrC)	NtrC	1-157 (REC domain)	NcoI, HindIII	471
$1633_{321\text{-}756}$ (NtrY)	NtrY	321-756 (PAS, HK core)	NcoI, HindIII	1315
$1634_{1\text{-}130}$ (NtrX)	NtrX	1-130 (REC domain)	NcoI, HindIII	390
$1656_{109\text{-}384}$	22475	109-384 (HAMP, HK core)	NdeI, HindIII	975
1657	22474	1-232 (full-length)	NdeI, HindIII	845
1929 (Hpt)	Hpt	1-122 (full-length)	NcoI, EcoRI	375
$2935_{321\text{-}614}$	DctB	321-614 (HK core)	NcoI, HindIII	890
$2936_{1\text{-}140}$	DctD	1-140 (REC domain)	NcoI, HindIII	425
$3039_{211\text{-}453}$	20047	211-453 (HK core)	NcoI, HindIII	735
3038	20048	1-223 (full-length)	NcoI, HindIII	675
3236	20899	1-413 (full-length)	NcoI, HindIII	1242
3235	20898	1-243 (full-length)	NdeI, HindIII	875
3733 (BphP1)	Phyto	1-703 (full-length)	NcoI, HindIII	2112
$3733_{462\text{-}703}$ (BphP1-HK)	PhytoHK	462-703 (HK core)	NcoI, HindIII	730
3734 (AgR)	AgR	1-152 (full-length)	NcoI, HindIII	460
3735 (ExsG)	ExsG	1-336 (full-length)	NdeI, HindIII	1155
$3735_{1\text{-}129}$ (ExsG-REC)	ExsG-REC	1-129 (REC domain)	NdeI, HindIII	530
$3735_{130\text{-}336}$ (ExsG-HK)	ExsG-HK	130-336 (HK core)	NdeI, HindIII	770
3735-H151N (ExsG-H151N)	ExsG H151N	1-336 (full-length), His151Asn	NdeI, HindIII	1155
3735-H192N (ExsG-H192N)	ExsG H192N	1-336 (full-length), His192Asn	NdeI, HindIII	1155
3735-D57N (ExsG-D57N)	ExsG D57N	1-336 (full-length), Asp57Asn	NdeI, HindIII	1155
3735-D62N (ExsG-D62N)	ExsG D62N	1-336 (full-length), Asp62Asn	NdeI, HindIII	1155
3735-D67N (ExsG-D67N)	ExsG D67N	1-336 (full-length), Asp67Asn	NdeI, HindIII	1155
3736 (ExsF)	ExsF	1-124 (full-length)	NcoI, HindIII	380
$3863_{301\text{-}594}$ (ChvG)	ChvG	301-594 (HK core)	NcoI, HindIII	890
3864 (ChvI)	ChvI	1-240 (full-length)	NcoI, HindIII	725
$4001_{191\text{-}462}$	90142	191-462 (HK core)	NcoI, HindIII	820
4002	90143	1-225 (full-length)	NcoI, HindIII	680

Table 2.2: List of all plasmid constructs. RE sites for diagnostic digestion and resulting fragment size are given.

Two µl of β-ME provided with the kit were added to 45 µl aliquots of thawed XL-10 Gold^{TM} Ultracompetent Cells in 14 ml BD Falcon round-bottom tubes and the cells were incubated on ice for 2 min. After adding 2 µl of the DpnI-treated DNA to separate aliquots, the cells were incubated on ice for 15-30 min, heat-shocked for 30 seconds at 42°C and incubated on ice for 2 min. To each tube 0.5 ml of pre-warmed LB media was added and the cells were grown at 37°C for 1 hour with shaking. Two volumes (80 and 180 µl) of the cells were plated onto separate LB agar plates containing 50 µg/ml kanamycin and the plates were incubated at 37°C overnight. Two bacterial colonies were picked per each successful transformation with mutagenised plasmid DNA and used to inoculate overnight cultures. The following day plasmid DNA was extracted and sent for sequencing (MWG Operon) to verify the success of mutagenesis.

CONSTRUCT	FORWARD PRIMER	REVERSE PRIMER
H151N	GGTCCGCGAGCTCAGC**A**ATCGCGTGAAGAACACC	GGTGTTCTTCACGCGAT**T**GCTGAGCTCGCGGACC
H192N	GCTTTGGCCGATGCG**A**ATGCCCTCCTGTTCG	CGAACAGGAGGGCAT**T**CGCATCGGCCAAAGC
D57N	GACGGGGGCGACTAT**A**ACGTCATCCTCGCAG	CTGCGAGGATGACGT**T**ATAGTCGCCCCCGTC
D62N	GACGTCATCCTCGCA**A**ACTTTTCCCTGCCCG	CGGGCAGGGAAAAGT**T**TGCGAGGATGACGTC
D67N	GACTTTTCCCTGCCC**A**ACTTCGACGGCATGAG	CTCATGCCGTCGAAGT**T**GGGCAGGGAAAAGTC

Table 2.3: Mutagenic primers used in site-directed mutagenesis of ExsG sequence. The mutated residue is shown in bold. The melting temperature of each primer was above 75°C.

2.2.17 Glycerol stock preparation

1 ml of overnight cultures of successfully transformed DH5α and BL21 (DE3) cells was mixed with 150 μl of glycerol; all glycerol stocks were stored at -80° C in cryogenic tubes.

2.3 Polyacrylamide gel electrophoresis (PAGE)

2.3.1 Sodium dodecyl sulphate (SDS)-PAGE

The gels used were prepared in-house, using Protogel (National Diagnostics) which contains 30% (w/v) acrylamide and 0.8% (w/v) bis-acrylamide. Stacking gel contained 4% and resolving gel 15% acrylamide, unless stated otherwise. The buffer used for

making the resolving gel contained 1.5 M Tris-Cl pH 8.8 and 0.4% (w/v) SDS, while stacking gel was prepared with 0.5 M Tris-Cl pH 6.8, 0.4% SDS buffer. Each 10 ml of stacking/resolving gel contained: 2.5 ml appropriate SDS buffer, 100 μl 1% ammonium persulfate solution, an appropriate amount of Protogel, water and 10 μl of TEMED.

Protein samples, diluted if necessary, were mixed with 4x loading dye (NuPage LDS Sample buffer, Invitrogen, mixed with 30 μl β-ME), heated at 99°C for at least 5 min and briefly centrifuged before loading. Running buffer contained 25 mM Tris, 192 mM glycine and 0.1% (w/v) SDS. Gels were electrophoresed at 150 V until the samples reached the resolving gel, then at 200 V for approximately 40 min. Staining normally involved 1 hour incubation in InstantBlue, a Coomassie-based instant staining solution supplied by Expedeon.

2.3.2 Phosphoprotein affinity-PAGE

The resolving gel (7.5 - 10%) was co-polymerised in the presence of 100 µM Phos-tag™ acrylamide and a twice higher concentration of $MnCl_2$. Stacking gel preparation and the running procedure were as specified for traditional SDS PAGE. The gels were run for an additional 10 - 15 min after the dye front disappeared from the bottom of the gel. The gels were first stained using InstantBlue (Expedeon) and then stained additionally with a solution containing Coomassie dye (0.25% (w/v) Coomassie Brilliant Blue, 40% (v/v) methanol, 7% (v/v) acetic acid and 52% (v/v) water) and kept for 30 min on a gel shaker. Destaining solution (40% (v/v) methanol, 10% (v/v) acetic acid and 50% water) was applied and removed several times until bands were clearly visible.

Whenever WT ExsG was being analysed, 7.5% polyacrylamide gels were used, while 10 % gels were used for the analysis of AgR, ExsG-REC and ExsF.

2.4 Protein expression and purification

2.4.1 Dialysis

Dialysis tubing, MWCO 12 - 14 kDa (Visking) or Spectra/Por 7 Dialysis Tubing, MWCO 3.5 kDa (Spectrum Laboratories) was washed with ultrapure water and with the appropriate dialysis buffer prior to dialysis. All dialyses involved overnight incubation at 4° C with stirring. After dialysis, protein solutions were filtered using 0.22 µm syringe filter.

2.4.2 Protein concentration and filtration

Proteins were concentrated using Vivaspin concentrators (GE Healthcare) of appropriate molecular weight cut-off. Small solution volumes (<2 ml) were filtered using Ultrafree-MC centrifugal filter units with Durapore membranes (Millipore) while larger volumes were filtered with Millex 0.22 μm syringe filters (Millipore).

2.4.3 Buffer exchange

Buffers were exchanged either using Vivaspin concentrators (GE Healthcare), with the desired buffer added >3 times in a volume at least 5 times greater than that of the protein sample, or with Micro Bio-Spin chromatography columns (Bio-Rad) according to manufacturer's instructions.

2.4.4 Protein concentration measurements

Protein concentrations were determined using Thermo Spectronic BioMate 3 spectrophotometer. Solutions were applied to a 1 ml quartz cuvette and the absorption was measured at 280 nm. Molar extinction coefficients for each protein were obtained from its amino acid sequence using the ExPaSy ProtParam online tool[8]. The concentration of a protein can be determined using the Beer-Lambert law:

$$A_{280} = \varepsilon l c$$

where A_{280}- protein absorbance at 280 nm, ε- molar extinction coefficient at 280 nm $[\frac{M}{cm}]$, c - protein concentration $[M]$, l - path length $[cm]$.

2.4.5 Tobacco etch virus (TEV) protease expression and purification

The recombinant TEV protease used in this work contained an N-terminal uncleavable His-tag and a C-terminal truncation was engineered in-house to render a more soluble protein which could be more expressed and purified at a significantly higher level than the wild-type.

Glycerol stock of *E. coli* BL21 (DE3) cells containing the plasmid with the mutant TEV protease gene was used to inoculate 50 - 100 ml of LB media with ampicillin and the culture was incubated overnight with shaking at 37° C. The following day 10 ml of

[8]http://web.expasy.org/protparam/

the culture was added per 1 L of ampicillin-containing LB media. The large cultures were grown at 37° C until the optical density at 600 nm reached at least 0.6 and induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (AppliChem) to a final concentration of 1 mM. The temperature was then lowered to 22° C and the cells were incubated for further 4 - 5 hours, and finally harvested at 2500 - 3000 x g.

The pellets were resuspended in the lysis buffer (25 mM Tris-Cl pH 8.2, 500 mM NaCl, 2 mM β-ME), sonicated using the thick probe (16 cycles, 10 sec on, 50 sec off, 10 mA amplitude) and centrifuged at 35000 x g. The lysate was incubated on a roller-shaker at 4° C for 1 hour with an appropriate volume (1 ml per 2L cell pellet) of Ni^{2+}-nitrilo-triacetic acid (NiNTA) chelating beads (Qiagen) pre-equilibrated in the lysis buffer. The beads were then centrifuged at 1000 x g and washed 5 - 6 times with ∼50 ml of the wash buffer (lysis buffer with 20 mM imidazole). The first elution involved 15 min incubation on a roller-shaker with 10 ml of the elution buffer (lysis buffer with 250 mM imidazole), 5 ml of the buffer was used for the second elution. The protein was dialysed overnight against 20 mM NaH_2PO_4 pH 5.6, 250 mM NaCl, 20% glycerol, 10 mM β-ME buffer, aliquoted and stored at -80° C. SDS-PAGE was used to determine the purity and yield of TEV protease.

2.4.6 Protein expression trials

BL21 (DE3) glycerol stocks were used to inoculate 2 ml of kanamycin-containing LB media and the cell cultures were incubated overnight at 37 ° C with shaking. The next day 1 ml of the culture was added to 9 ml of fresh LB media with kanamycin and, following incubation for further 40 min, expression was induced by adding IPTG to a final concentration of 1 mM. Two hours post-induction, the cells were harvested, resuspended in approximately 5 ml of lysis buffer (0.5 M NaCl, 50 mM Tris-Cl pH 8.2, 1 mM β-ME and protease inhibitors) and sonicated for 90 seconds. Lysates were centrifuged at 35000 x g for 25 min and the supernatant was incubated for one hour on a roller shaker with 100 μl of NiNTA beads (Qiagen) (equilibrated in the lysis buffer). After spinning for 5 min at 1000 x g, the supernatant was removed; approximately 15 ml of wash buffer (50 mM Tris-Cl pH 8.2, 0.5 M NaCl, 20 mM imidazole, 1 mM β-ME) was added and the tube was centrifuged again. Four to five wash buffer exchanges were made, then 200 μl of elution buffer (50 mM Tris-Cl pH 8.2, 0.5 M NaCl, 300 mM imidazole, 1 mM β-ME) were added and the sample was incubated on a roller shaker for 10 min after which the protein sample was separated by centrifugation. SDS-PAGE was used to analyse 15 μl of each pellet, flow-through, elution as well as beads following elution with imidazole.

2.4.7 Large-scale protein expression

LB media supplemented with kanamycin were inoculated (10-20 ml per litre of media) with the overnight culture which was prepared the previous night using the appropriate glycerol stock. The cells were incubated at 37° C with shaking until the OD at 600 nm was at least 0.6, induced with IPTG to give a final concentration of 1 mM, grown for approximately 3 hours at 30° C (ExsG and ExsG-HK) or 23-25° C (AgR, ExsF, BphP1-HK, ExsG-REC) and harvested at 2500 - 3000 x g.

ArcticExpress™ cells after inoculation were grown with shaking at 37° C until optical density reached 0.3 - 04 and the temperature was lowered to 30° C. When optical density was approximately 0.7, the cells were chilled to 14° C and induced with 0.4 mM IPTG. They were then grown at 14° C for ∼24 hours and harvested.

2.4.8 Purification with NiNTA beads

All cell pellets (except for the cells expressing Hpt domain protein) were resuspended in 20 - 30 ml (3 - 5 ml per gram of cells) of lysis buffer (50 mM Tris-Cl pH 8.2, 0.5 M NaCl, 1 mM β-ME) containing 1 cOmplete protease inhibitor cocktail tablet (Roche) and the suspensions were sonicated using the thick probe (16 cycles, 10 sec on, 50 sec off, amplitude 10 mA). The lysate was centrifuged at 35000 x g for 30 min. The supernatant was incubated with an appropriate volume of pre-equilibrated NiNTA beads (Qiagen) on a roller shaker for 1 hour. The beads were washed 5 - 7 times with ∼50 ml of the wash buffer (lysis buffer with 20 mM imidazole). The protein was typically eluted with 10 ml (first elution) and 5 ml (second elution) of the elution buffer (lysis buffer with 300 mM imidazole). The purity of the elutions was determined using SDS-PAGE gel. BphP1-HK purification buffers contained also 5% (v/v) glycerol.

Hpt domain protein purification was initially attempted using the same method as above (with 0.5 - 1 ml NiNTA beads per 6 g of cells) but the buffers used contained 50 mM MES (2-(N-morpholino)-ethanesulfonic acid) pH 6, 1 M NaCl, 1 mM β-ME and 15% (v/v) glycerol. High salt concentration was used as the protein appeared to be highly charged, while low pH was chosen due to the basic isolelectric point (7.2 with and 8.2 without the purification tag). After tag removal the protein was dialysed against 50 mM Tris-Cl pH 7, 0.5 M NaCl, 1 mM β-ME and 15% (v/v) glycerol buffer. The other purification method involved buffer containing 50 mM Tris-Cl pH 8.5, 0.5 M NaCl, 1 mM β-ME and 10% glycerol and the purification tag was not cleaved; instead, the eluted protein was dialysed against the purification buffer without imidazole.

2.4.9 Removal of purification tag with TEV protease

An appropriate amount of TEV protease (1 mg of TEV per 10 - 15 mg of protein) was added to protein eluted from NiNTA beads. Protein-TEV solution was dialysed against 2 litres of 50 mM Tris-Cl pH 8.2, 250 mM NaCl, 1 mM β-ME buffer overnight. The next day an appropriate volume of equilibrated NiNTA (100 - 200 µl per 1 mg of TEV protease used) was added and the solution was incubated on a roller shaker for 40 min and then centrifuged for 5 min at 1000 x g.

2.4.10 Size-exclusion chromatography (SEC)

Concentrated protein samples (minimum 1 mg of protein per injection) were injected onto pre-packaged Superdex 200 10/300, Superdex 75 10/300 or Superose 6 10/300 columns (GE Healthcare), which were previously washed with water and pre-equilibrated with the appropriate degassed and filtered buffer. The standard SEC buffer contained 50 mM Tris-Cl pH 8.2, 1 mM β-ME and 30-500 mM NaCl; 5% (w/v) glycerol was also included during BphP1-HK purification. The buffer used in SEC of proteins for electrospray ionisation mass spectrometry contained 150 mM ammonium acetate pH 8.2, 100 µM dithiothreitol (DTT). SEC was performed at temperature below 10° C using the ÄKTA FPLC system, with the flow rate 0.5 ml/min. Elution fractions containing UV absorption peaks on the chromatogram were examined by SDS-PAGE and those with the pure protein were combined. The columns were calibrated using high or low molecular weight markers (HMW and LMW Gel Filtration Calibration Kit, GE Healthcare) in the appropriate buffer to allow estimation of the protein sizes. For molecular weight determination analytical pre-packaged Superdex 75 PC 3.2/30 column (GE Healthcare) was used with the flow rate of 0.04 ml/min; 0.05 - 0.1 mg of protein was injected per run.

2.5 Preparation of nucleotides, phosphate donors and beryllofluoride

2.5.1 Nucleotides

2.5.1.1 Preparation of stocks

ATP and its two non-hydrolysable ATP analogues were used throughout this project - ADP and AMP-PNP (Sigma). The nucleotide stocks were prepared by dissolving the powder either in water (radioactive assays) or in an ammonium acetate solution (mass

spectrometry) and the concentration was determined by measuring the absorbance at 259 nm (extinction coefficient = 15.4 $mM^{-1}cm^{-1}$). The stocks were always prepared freshly on the day of the experiment and kept at 4° C.

2.5.1.2 Ethanol precipitation of AMP-PNP

Stock solution of 50 mM AMP-PNP was prepared in ∼2.5 M ammonium acetate, mixed with 2 volumes of ethanol and incubated for 2 hours. The mixture was centrifuged at 17000 rpm using a bench-top microfuge and the supernatant was decanted. The pellet was washed with 70% ethanol, spun briefly and the supernatant was removed. The pellet was air-dried and resuspended using ammonium acetate to give a final concentration of 2.5 M. The procedure was repeated twice and the final pellet was resuspended in the buffer used in mass spectrometry experiments (150 mM ammonium acetate pH 8.2, 100 μM DTT). The concentration of the purified AMP-PNP was determined based on the absorbance at 259 nm.

2.5.2 Berylloflouride

Berylloflouride was always freshly prepared by mixing stock solutions of 0.2 M beryllium chloride ($BeCl_2$, Sigma) with approximately 6-fold molar excess of 0.8 M sodium fluoride (NaF, Sigma); $BeCl_2$ stock solution was kept at 4° C for up to 1 month while NaF solution was prepared fresh.

2.5.3 Acetyl phosphate

Lithium potassium acetyl phosphate (Sigma) was prepared fresh on the day of experiment by dissolving the powder in water, to a final concentration of 0.5 M.

2.5.4 Phosphoramidate (PA)

Potassium hydrogen phosphoramidate is not commercially available and was therefore synthesised in-house using two steps from the standard protocol established by Sheridan and co-workers [123] (see below). The obtained powder was dissolved in water to a final concentration of 0.5 M on the day of experiment and the solution was kept at 4° C.

2.5.4.1 Ammonium hydrogen phosphoramidate synthesis

18.3 ml of phosphoryl chloride ($POCl_3$, Sigma) was added dropwise over the course of 5-10 min to 300 ml of an ice-cold 10% solution of ammonia (Sigma) with constant stirring.

The solution was stirred for further 15 min and 1 litre of acetone was added which resulted in separation into two layers. The bottom layer was separated, neutralised with 8 ml of glacial acetic acid to pH ∼6 and cooled to temperature below 10° C to induce PA crystallisation. A volume of ethanol equal to the volume of the filtrate was added to promote further crystallisation and the salt was filtered, washed successively with ethanol and diethyl ether, and allowed to air dry. The yield was ∼12 grams.

2.5.4.2 Potassium hydrogen phosphoramidate synthesis

Potassium hydrogen phosphoramidate was prepared by dissolving 11.4 g of ammonium hydrogen phosphoramidate in 20 ml of 50% solution of potassium hydroxide and heating at 50-60° C for 10 min to evaporate ammonia. Afterwards the solution was cooled to below 10° C, neutralised to pH 6 with glacial acetic acid and mixed with 1 litre of ethanol which promoted potassium salt precipitation. The crystals were filtered, washed successively with ethanol and diethyl ether, and allowed to air dry. The yield was 7.9 grams.

2.6 Phosphorylation and phosphatase assays

2.6.1 Phosphoprotein affinity PAGE

All reactions were performed in 50 mM Tris-Cl pH 8.2, 250 mM NaCl, 1 mM β-ME buffer (unless stated otherwise) at room temperature and stopped by adding the appropriate loading dye. Micro Bio-Spin chromatography columns with bio-gel in Tris-Cl pH 7 buffer (Bio-Rad) were prepared or pre-equilibrated as instructed in the manual provided. The extent of phosphorylation was proportional to the intensity of up-shifted protein bands on the gel. This intensity was quantified where necessary using GelQuant Express (DNR Bio Imaging Systems) with the background set manually for each set of reactions.

2.6.1.1 ExsG receiver domain phosphorylation

Purified WT ExsG (30 μM) and ExsG-D62N (40 μM) were allowed to autophosphorylate with PA (50 mM) in the presence of 50 mM $MgCl_2$ for 2 hours with WT ExsG samples taken at regular intervals and ExsG-D62N sample taken after 2 hours. In a negative control experiment separate WT and D62N samples were simultaneously incubated without the PA.

2.6.1.2 RR autodephosphorylation assay

For measuring autodephosphorylation rates ExsF, AgR, WT ExsG and ExsG-REC (∼50-70 μM each) were allowed to autophosphorylate with 50 mM PA in the presence of 50 mM $MgCl_2$ for 2 hours (WT ExsG) or 1 hour (AgR, ExsF and ExsG-REC) and excess of the compound was removed by passing the proteins through the chromatography spin columns. The proteins were then supplemented with magnesium chloride to a final concentration of 10 mM and samples were taken at several time points to assess the progress of dephosphorylation by gel electrophoresis. ExsF was assayed separately following 1.5 hour phosphorylation with 50 mM acetyl phosphate in the presence of 50 mM $MgCl_2$.

2.6.1.3 BphP1-HK phosphatase assay

WT ExsG was incubated with 50 mM PA and 50 mM $MgCl_2$for 2 hours after which unreacted compound was removed. BphP1-HK, $MgCl_2$ and ADP were added to a final concentration of 30 μM, 10 mM and 1 mM, respectively. The final concentration of ExsG in the reaction mixture was about 30 μM. The proteins were incubated over 2 hours with samples taken at specific time points and examined by gel electrophoresis.

2.6.2 Radioactive assays

γ-[^{32}P]-ATP (6000 Ci/mmol) was purchased from Perkin Elmer. The "hot" reaction buffer, unless stated otherwise, contained 50 mM Tris-Cl pH 8.2, 250 mM NaCl, 1 mM β-ME, 10 mM $MgCl_2$, 7 mM ATP and radioactively labelled ATP such that each sample loaded on the gel would contain 5-10 μCi. In the reaction the final concentration of ATP was 5 mM and of $MgCl_2$ - approximately 7 mM. All reactions were performed at room temperature and stopped either by addition of the loading dye (NuPage, Invitrogen) or by elution from Bio-spin chromatography columns in Tris-Cl pH 7 buffer (Bio-Rad) which would remove ATP or PA from the protein solution. Where necessary, the spin columns were pre-equilibrated into 50 mM Tris-Cl pH 8.2, 250 mM NaCl, 1 mM β-ME, 10 mM $MgCl_2$ buffer.

Proteins were separated either by standard SDS-PAGE (15% acrylamide) or phosphoprotein affinity gel electrophoresis; the gels were wrapped in cling-film and placed into a cassette with a phosphorimaging screen. After 2 hours the screen was scanned using a phosphorimager (Fuji FLA-2000) and the gels were stained appropriately to estimate protein levels from band intensities. The intensity of the radioactive signal was

quantified using the programme MultiGauge v. 2.0 (Fuji Film) with the background set manually.

2.6.2.1 Autokinase assays

For the assessment of autokinase activity WT ExsG and its variants (40 μM), ExsG-HK ($\sim$35 μM) and BphP1-HK ($\sim$20 μM) were allowed to autophosphorylate in the "hot" reaction buffer (note: final ATP concentration in the reactions was 1 mM). Samples were taken at regular intervals and stopped by addition of SDS to give a final concentration of 2% (w/v). The excess of nucleotide was removed using chromatography spin columns prior to mixing with the protein loading dye, electrophoresis and autoradiography. To test for the autokinase activity in ammonium acetate buffer, the protein (40 μM) was incubated in a "hot" buffer containing 150 mM ammonium acetate pH 8, 100 μM DTT, 10 mM $MgCl_2$, 5 mM ATP for 1 hour.

2.6.2.2 ATP dependence assays

Optimal ATP concentration for the autokinase reaction was estimated by mixing BphP1-HK (15 μM final concentration) or WT ExsG (30 μM) aliquots with the "hot" reaction buffer containing varying final ATP concentrations (0.01, 0.05, 0.2, 0.5 and 1 mM) for one hour. Subsequently, the reaction mixtures were cleaned from excess ATP by chromatography spin columns and analysed using SDS-PAGE and autoradiography.

2.6.2.3 Autokinase assays on (Asp$\sim$P)-ExsG

To determine the effect of phosphorylation of the receiver domain on autokinase activity, two different experiments were performed. WT ExsG (30 μM) was incubated for 1 hour in the "hot" reaction buffer in the presence of different concentrations of beryllofluoride. $BeCl_2$ at 0.2, 0.5, 1, 3, and 5 mM concentrations with 6-fold molar excess of NaF in each case were used. A control reaction without the compound was incubated alongside. Prior to loading on the gel, the proteins were passed through the chromatography spin columns.

In the second experiment WT ExsG and ExsG-D62N (120 μM each) were incubated for 2 hours in the presence of 50 mM $MgCl_2$ with or without 50 mM PA. All samples were passed through the chromatography spin columns and an appropriate volume of the "hot" reaction buffer was added such that the final concentration of the proteins was approximately 30 μM. Samples were taken out of each reaction at specific time points and analysed by SDS-PAGE and autoradiography.

2.6.2.4 Assay for the acid/base stability

BphP1-HK (30 μM) and WT ExsG (40 μM) were allowed to autophosphorylate in hot reaction buffer for 1 hour, passed through the chromatography spin columns and three samples of each protein were loaded on the gel. After determination of the phosphorylation level by autoradiography, the gel was covered with a gel-size strip of Hybond^{TM} membrane (GE Healthcare) which was pre-equilibrated first in methanol and subsequently in transfer buffer (100 mM Tris-Cl pH 8, 150 mM NaCl, 0.03% SDS, 2% methanol). Three gel-size strips of blotting paper were placed on both gel and membrane sides and the proteins were transferred onto the membrane by electroblotting in transfer buffer for 2.5 hours at constant current of 250 mA. Positions of the protein bands were visualised by 5 min staining of the membrane with Ponceau S stain (0.1% Ponceau S, Sigma, in 5% acetic acid solution). The membrane was cut into strips, each containing one BphP1-HK or ExsG band, and each strip was kept for 2 hours in one of the following solutions: 50 mM Tris-Cl pH 8.2, 1 M HCl or 3 M KOH. Afterwards the strips were assembled together and subjected to autoradiography.

2.6.2.5 Phosphohistidine stability assays

BphP1-HK (15 μM) and WT ExsG (30 μM) were allowed to autophosphorylate in hot reaction buffer for 90 min. After removal of excess nucleotide samples were taken at specific time points and analysed by SDS-PAGE and autoradiography.

2.6.2.6 Transphosphorylation assays and phosphotransfer profiling

The HKs (ExsG or BphP1) were autophosphorylated as outlined before for 30 (BphP1-HK) or 60 min (ExsG). RRs (ExsF, ExsG-REC, AgR, WT ExsG and ExsG-H151N) were added to autophosphorylated HK proteins such that they were present in approximately 3-fold molar excess, except for phosphotransfer profiling using BphP1-HK whose concentration was at least 6-fold lower than that of the RRs. Phosphotransfer from BphP1-HK onto ExsG was monitored using three time-points: 5, 30 and 60 min. In phosphotransfer profiling the reactions were incubated for 5 and 60 min, stopped by addition of SDS to a final concentration of 2% and unreacted ATP was removed prior to addition of the gel loading dye. In transphosphorylation assay which served to identify the phosphorylatable aspartate residue of ExsG the reactions were incubated for 1 hour. For the preferential assay, phosphorylated BphP1-HK (10 μM, 1 hour autophosphorylation) was incubated with 40 μM AgR and ExsG-H161N for 5 and 60 min.

In experiments testing the ability of ExsG to act as a RR the excess of nucleotide was removed from BphP1-HK autophosphorylation reaction prior to addition of WT ExsG and ExsG site-specific variants. Subsequently an appropriate amount of $MgCl_2$ was added to a final concentration of 10 mM. The exceptions were the phosphotransfer profiling and the preferential assay where the inactive variant ExsG-H151N was used.

2.6.2.7 Phosphatase assays

BphP1-HK (10 μM) was allowed to autophosphorylate for 1 hour and the excess of nucleotide was removed. It was then incubated for 1 hour with WT ExsG or ExsG-D67N (100 μM each); $MgCl_2$ was added to each reaction at a final concentration of 10 mM. The samples were then incubated in the presence or absence of 0.5 M ADP for 5, 90 and 180 min.

ExsG (15 μM) was autophosphorylated for 1 hour in the presence of ExsF (100 μM) and the unincorporated nucleotide was removed. $MgCl_2$ was added to a final concentration of 10 mM and the samples were incubated in the presence or absence of 0.5 M ADP for 1, 15, 30, 60 and 120 min.

2.6.2.8 Hpt domain protein phosphotransfer tests

Hpt protein purified at pH 6 and dialysed into Tris-Cl pH 7 buffer was concentrated to $\sim$50 μM. WT ExsG (50 μM) was incubated with ExsF (80 μM) and BphP1-HK (40 μM) was incubated with AgR (70 μM) in the "hot" buffer for 90 min after which the reactions were passed through the chromatography spin columns pre-equilibrated with 50 mM Tris-Cl pH 7, 0.5 M NaCl, 1 mM β-ME buffer. To 5 μl of each elution 15 μl of Hpt protein or 50 mM Tris-Cl pH 7, 0.5 M NaCl buffer was added and the reactions were incubated for further 2 hours prior to loading on the gel.

Hpt protein purified at pH 8.5 with the purification tag retained was concentrated to $\sim$1 mg/ml. Transphosphorylation of the RR protein was carried out by incubating 30 μM BphP1-HK with 80 μM AgR and 60μM ExsG with 90 μM ExsF separately in the "hot" buffer for 1 hour after which 15 μl of Hpt was added to each transphosphorylation reaction. NiNTA beads (15 μl) were pre-equilibrated in 50 mM Tris-Cl pH 8.5, 0.5 M NaCl, 1 mM β-ME buffer and added to separate aliquots of Hpt-containing reactions, either at the beginning or after 1 hour of the incubation. In reactions containing beads throughout the incubation, the tubes were gently agitated every few minutes. Tubes to which the beads were added after the incubation were agitated several times over the period of 10 min. All tubes were briefly centrifuged at 1000 x g and both elution

and the beads fractions were loaded onto the gel. Prior to loading the samples onto the gel, 10 μl of the pH 8.5 buffer was added to the beads fraction to resuspend them. The samples were analysed by SDS-PAGE and autoradiography.

2.6.2.9 Test of beryllofluoride effect on ExsG autophosphorylation and phosphohistidine stability

Stability of a phosphohistidine was assayed by incubating 40 μM WT ExsG and 15 μM ExsG-HK in the "hot" buffer for over an hour. After the removal of nucleotides by passing the reaction mixture through the chromatography spin columns beryllium chloride and sodium fluoride were added to half of each elution to a final concentration of 5 mM $BeCl_2$ and 30 mM NaF while water was added to the remaining parts of the elutions to provide control measurements. Samples were taken at specified time points and analysed using SDS-PAGE and autoradiography.

Putative effect of beryllofluoride on the rates of autophosphorylation was assayed as follows: WT ExsG and ExsG-HK were incubated in hot reaction buffer for over an hour in the absence or presence of beryllofluoride. Two $BeCl_2$ /NaF concentrations were used (0.2/1.2 and 5/30 mM) for WT ExsG reactions and three (0.2/1.2, 1/6 and 5/30) for ExsG-HK reactions. Afterwards, the reactions were analysed by SDS-PAGE and autoradiography.

2.6.2.10 Test of the effect of phosphorylation of ExsG receiver domain on ExsG autophosphorylation and phosphohistidine stability

For the autophosphorylation assay samples of WT ExsG and ExsG-D62N ($\sim$30 μM each) were incubated separately with or without 50 mM PA/$MgCl_2$ for 2 hours. After the compound was removed from the relevant reactions using chromatography spin columns, the proteins were incubated in the "hot" buffer for 1 hour. For the phosphohistidine stability assay, WT ExsG ($\sim$40 μM) was allowed to autophosphorylate for 1 hour and after removing unincorporated nucleotide the protein was incubated in the presence of 50 mM PA/$MgCl_2$ for 1 hour. Samples taken at specified time points from each reaction were analysed using SDS-PAGE and autoradiography.

2.6.2.11 Radioactive phosphoprotein affinity gel electrophoresis - double phosphorylation experiments

WT ExsG, BphP1-HK and ExsG-HK (30 - 40 μM each) were used for both non-radioactive as well as radioactive phosphoprotein affinity gel electrophoresis which ser-

ved to differentiate between the mobility shifts of (Asp-∼P)- and (His∼P)-ExsG. In autokinase assays all proteins were incubated with the non-radioactive or "hot" reaction buffer for approximately one hour. Prior to the transphosphorylation reaction, BphP1-HK was allowed to autophosphorylate for 30 min and excess ATP was removed. The reaction, containing ∼10 μM BphP1-HK∼P and 40 μM ExsG, was incubated for 1 hour.

Two types of double phosphorylation experiments were performed on WT ExsG. First one involved phosphorylation of the receiver domain of ExsG (60 μM) with 50 mM PA/$MgCl_2$ over varying periods of time, removal of excess compound and incubation of the protein in the "hot" buffer. The progress of the autokinase reaction was monitored over 1 hour. A control reaction was set up alongside where ExsG was incubated without PA prior to the autokinase reaction.

In the other type of experiment 40 μM ExsG was allowed to autophosphorylate in the "hot" buffer for 90 min, the unincorporated nucleotide was removed and PA and $MgCl_2$ were added to a final concentration of 50 mM. Samples were taken before addition of PA and after, at 5, 45 and 90 min.

2.7 Protein crystallisation trials

2.7.1 Manual set-up

ExsG full-length (5-10 mg/ml), ExsG-HK (4-6 mg/ml) and ExsF (6-12 mg/ml) were initially subjected to crystallisation trials using hanging drop vapour diffusion. All proteins were tested to determine favourable crystallisation conditions using screens HR2-110 and HR2-112 (Hampton Research). ExsG was also tested with Structure Screen I and II (Molecular Dimensions). Each drop was prepared by mixing 1 µl of the protein with 1 µl reservoir buffer and equilibrated against 0.5 ml reservoir solution; during each trial several different concentrations of the proteins were used. The plates were incubated at 20° C and checked every day for the first week, then 2-4 times per month.

2.7.2 Automated set-up

In order to test a larger array of crystallisation conditions, each protein was subjected to trials with six crystallisation screens: JCSG, PACT, PEG/SALT, Classic I and Classic II (Jena Bioscience) and Index (Hampton Research). The plates were set up using Mosquito nano-litre liquid handling robot (TTP Labtech). Each sitting drop contained

0.1 μl of the protein and an equal volume of reservoir solution; 80 μl of the precipitating reagent was dispensed in the reservoir. The trays were incubated at 16^o C and examined on a regular basis using the Minstrel DT UV imaging system (Rigaku).

2.7.3 Cofactors and phosphate/ATP analogues

Crystallisation of ExsG full-length and ExsG-HK in the presence of cofactors was attempted manually and in the automated fashion with addition of magnesium chloride (2-5 mM), ATP analogues AMP-PNP and ADP (Sigma) (0.5-1 mM), as well as with both magnesium chloride and an ATP analogue. ExsG full-length and ExsF were also subjected to manual crystallisation trials in the presence of the phosphate-mimic beryllofluoride (5 mM $BeCl_2$ and 30 mM NaF) and 10 mM magnesium chloride. The cofactors and analogues were added to the protein solution and the mixtures were incubated for at least 10 min before dispensing the drops.

2.8 Analytical ultracentrifugation (AUC)

Densities of the buffers used in the AUC experiments were determined using a densitometer at 20^o C by Jayesh Gor (ISMB, UCL).
50 mM Tris-Cl pH 8.2, 30 mM NaCl - 1.001329 g/ml
50 mM Tris-Cl pH 8.2, 250 mM NaCl - 1.010859 g/ml
Partial specific volume of WT ExsG (0.7439 ml/g) was calculated using the programme SLUV (S. Perkins). All experiments were performed using Beckman An-50 Ti rotor, 8-place. Sedimentation velocity 12 mm column centrepiece cells have two sector-shaped channels, while those used in equilibrium experiments were six channel short-column (3 mm) centrepiece cells. Both experiments were conducted by Jayesh Gor.

2.8.1 Sedimentation velocity (SV) experiment

Seven different concentrations (0.2, 0.5, 0.7. 1, 1.25, 1.5 and 2 mg/ml) of WT ExsG were prepared in 50 mM Tris-Cl pH 8.2 buffer, with high (250 mM) and low (30 mM) NaCl concentrations. One of the two sector-shaped channels in SV centrepiece cells was filled with 420 μl of the appropriate buffer as a reference, while 400 μl of the sample was applied to the other channel. The cells were inserted into the rotor, aligned and an A280 scan was performed to confirm protein concentration in each cell. The experiment was run in a Beckman Optima L Series Ultracentrifuge at 20^o C. Initial rotor speed applied was 50000 rpm, but most of the protein sedimented to the bottom of the cell

within the first hour, and hence a lower speed - 40000 rpm - was used in the subsequent experiment. Interference data (concentration fringe migration in time) were collected by taking a scan every 1 min, producing 600 scans in total. After the experiment all samples were extracted and analysed by SDS-PAGE to ensure that the protein was not damaged by centrifugation. Fitting of parameters to the experimental data was performed in SEDFIT [124].

2.8.2 Sedimentation equilibrium (SE) experiment

Six concentrations (0.1, 0.2, 0.5, 0.7, 1 and 1.5 mg/ml) of the protein were prepared in 50 mM Tris-Cl pH 8.2 buffer with two different NaCl concentrations (30 and 250 mM). In each An-50 Ti rotor SE cell there are three pairs of compartments; one compartment of each pair was filled with 110 μl of the appropriate buffer, and the other with 100 μl of one of the samples. The experiment was run in Beckman ProteomeLab^{TM} XL-I Ultracentrifuge. Scans of absorbance at 280 nm were performed to confirm protein concentration in each compartment. Starting speed was 7000 rpm, followed by 11000 rpm, 16000 rpm, and finally 24000 rpm. Absorbance and interference scans were taken every 3 hours, and 10 scans were taken at each speed, giving the system 30 hours to reach equilibrium. After the run (one week), protein samples were analysed by SDS-PAGE together with corresponding samples which were left at room temperature for the same amount of time, to determine if centrifugation affected the state or degradation rate of the protein. Testing for approach to equilibrium and multi-speed file assembly was performed in SEDFIT [124], while single and global single- or multi-speed analyses were performed in SEDPHAT.

2.9 Mass spectrometry (MS) and ion mobility spectrometry (IMS)

All MS and IMS experiments were performed in collaboration with Dr Konstantinos Thalassinos (ISMB, UCL). ESI MS and IMS experiments and analysis were performed by Jun Yan.

2.9.1 Electrospray ionisation time-of-flight (ESI-ToF) MS

WT ExsG, ExsG-H161N and ExsF were concentrated to 30 - 40 μM (ExsG) or ~15 μM (ExsF) and buffer exchanged to a buffer composed of 150 mM ammonium acetate pH 8 and 100 μM DTT using Bio-Spin 6 chromatography spin columns (Bio-Rad). For the nucleotide binding experiment, ExsG was incubated with 1 mM AMP-PNP and 2 mM $MgCl_2$ at room temperature for 30 minutes, after which the incubation mixture was passed through the chromatography spin column. For the experiment involving phosphorylation of ExsG receiver domain, the protein was incubated with 50 mM $MgCl_2$ and 50 mM PA for 3 hours, and passed through two chromatography spin columns. Prior to the assembly pathway experiments, ExsG was buffer exchanged to buffers containing 300 mM, 400 mM, 600 mM, or 800 mM ammonium acetate, pH 8, 100 μM DTT using chromatography spin columns.

Mass spectrometry experiments were carried out on a first-generation Synapt HDMS (Waters Corporation) Quadrupole-ToF mass spectrometer [125]. 2.5 µl aliquots of the protein sample were introduced to the mass spectrometer by means of nano-electrospray ionisation using gold-coated capillaries prepared in-house. Typical instrumental parameters were as follows: source pressure 5.5 mbar, capillary voltage 1.2 kV, cone voltage 100 V, trap energy 20 V, transfer energy 10 V, bias voltage 22 V. Mass spectra were smoothed and peak-centred in MassLynx (Waters Corporation).

2.9.2 Ion mobility spectrometry - mass spectrometry (IMS-MS)

Typical instrumental parameters were as follows: IMS pressure 5.20 x 10-1 mbar, IMS wave velocity 250 m/s, IMS wave height 10 V, and trap pressure 4.07 x 10^{-2} mbar. Data analysis was performed in DriftScope (Waters Corporation).

2.9.3 Liquid-chromatography (LC)-tandem MS

Protein bands of interest were excised from polyacrylamide gels and the proteins were digested with trypsin according to the protocol by Shevchenko and co-workers [126]. Gel fragments were divided into small pieces and washed with 0.5 ml of 40% acetonitrile, 60% 50 mM ammonium bicarbonate (AmBic) for 30 min with shaking. The liquid was removed and gel pieces were dehydrated with 0.5 ml neat acetonitrile for 10 min until they became white. The liquid was removed and the pieces were dried at 50°C with open lids for a few minutes. Reduction involved adding 100 μl of 10 mM DTT and incubation at 80°C for 30 min. After cooling, 100 μl of 55 mM iodoacetamide was added and tubes were incubated in the dark at room temperature for 20 min to

allow alkylation. The liquid was removed and gel pieces were washed twice with 0.5 ml 50 mM AmBic, dehydrated with 200 μl neat acetonitrile and swelled with 200 μl AmBic; before each step the liquid from previous step was removed. Dehydration and swelling steps were repeated after which the pieces were dehydrated again, the liquid was removed and the pieces were dried briefly at 50°C. 50 mM AmBic containing 10 ng/μl porcine modified trypsin (Promega), prepared fresh, was added so that the pieces were covered. After 18 hours of digestion at 37°C an equal volume of extraction solution (1% formic acid, 2% acetonitrile) was added and the pieces were incubated for 30 min with occasional vortexing. The liquid was transferred to a fresh tube, pieces were dehydrated with 50 μl neat acetonitrile to extract hydrophobic peptides and the two extractions were combined. The extracted peptides were dried using a speed vacuum concentrator and resuspended in 0.2% formic acid, 2% acetonitrile solution.

LC-tandem MS experiment and analysis were conducted by Adam Cryar (UCL). Nano-RPLC separation of the tryptic digest samples was conducted using a nano-ACQUITY UPLC system (Waters Corporation) with a 5 µm SYMMETRY C18, 180 µm x 200 mm trap column and a 1.7 µm BEH130 C18, 75 µm x250 mM analytical column. A two phase linear gradient was applied where solvent A was 0.1 % formic acid in water and solvent B was 0.1 % formic acid in acetonitrile (J. T. Baker). Sample was applied at a flow rate of 5.0 µl/min to the trapping column at mobile phase composition of 3% solvent B. The column was then desalted for 4 minutes at the same conditions. A linear gradient was then applied over 29 min at a flow rate of 0.3 µl/min with the concentration of solvent B increasing from 3% to 40 %. MS analysis was conducted using a Synapt HDMS (Waters Corporation) coupled to the nano-ACQUITY system. Accurate mass measurements were made using a data-independent acquisition mode (LC-MS^{E}), whereby the energy in the collision cell was alternated from low energy (4 eV) to high energy (energy ramp from 15-35 eV) whilst continuously acquiring MS data. The Synapt time-of-flight mass analyser was calibrated over 175.11 - 1285.54 Da m/z range using fragment ions of the peptide [Glu1]-fibrinopeptide B (GFP). The doubly charged precursor monoisotopic peak of GFP was fragmented with a collision energy of 30 eV. GFP solution (500 fmol/µl) was introduced during the analysis at a flow rate of 0.5 µl/min via a NanoLockSpray source which facilitated the post-acquisition data lockmass correction using the monoisotopic mass of GFP. The reference sprayer was sampled every 60 s. Measurements were made over a m/z range of 100-2000 Da and the time-of-flight mass analyser was operated in V mode with a scan time of 0.6 seconds.

The data were processed using the programme PLGS v2.5 (Waters Corporation).

The computational methods implemented are explained in detail by Geromanos and co-workers [127]. Data were searched against an Sprot specific protein database appended with the sequence of ExsG. Carbamidomethyl-C was specified as a fixed modification. Oxidation (M), phosphorylation (S,T,Y) and acetylation (K, N-terminus) were specified as variable modifications. A maximum of two missed protease cleavages were allowed for semi-tryptic peptide identification. Two corresponding fragment ions were set as a minimum criterion for peptide identification, while for protein identification at least one corresponding peptide and seven fragment ions were required.

2.10 Isothermal titration calorimetry (ITC)

Purified WT ExsG was dialysed overnight against 20 mM Tris-Cl pH 8.2, 125 mM NaCl, 1 mM β-ME buffer with or without 2 mM $MgCl_2$ and the post-dialysis buffer was used to prepare the stock solutions of ADP. The protein was concentrated to 35, 80, 100 or 120 μM, filtered using a 0.22 μm syringe filter and degassed with stirring at 20°C using ThermoVac (GE Healthcare). ADP solutions used were of 10- or 16-fold higher concentration than the protein. ExsG-H161N (∼60 μM) was also tested in the same way with ATP as a ligand.

Control experiments consisted of ligand-into-buffer injections to determine the heats of dilution. The experiments were performed using VP-ITC machine (GE-Healthcare) with cell temperature set to 20°C (except for one experiment performed at 15°C) and 20 - 25 10 μl injections (first injection 5 μl). When stirring (300 rpm) was applied the spacings between injections were 150 seconds long; when stirring option was switched off, the spacing was extended to 400 seconds. After injecting the protein sample into the cell, the machine was allowed to equilibrate the cell temperature over at least 30 min before the start of the experiment.

2.11 Electron microscopy (EM) by negative staining

EM experiments were performed in collaboration with Prof Helen Saibil (Institute of Structural and Molecular Biology, Birkbeck College) with the assistance of Dr Luchun Wang. Thin carbon grids and 2% uranyl acetate stain were prepared in-house by Dr Natalya Lukoyanova, 2% phosphotungstic acid (PTA) solution was prepared by Michal Koenner.

WT ExsG samples were buffer exchanged into either in 50 mM Tris-Cl pH 8.2, 125 mM NaCl, 1 mM β-ME or in 150 mM ammonium acetate pH 8.2, 1 mM β-ME.

Several protein dilutions were prepared, ranging from 2 to 0.1 mg/ml. Carbon grids were ionised with a glow discharge for 1 min at 50 mA. 5 µl of the protein solution was placed on the carbon grid and incubated for 1 min, after which the liquid was blotted off with filter paper; a wash step with a 3 drops of water was sometimes incorporated before removing the liquid.

Staining with uranyl acetate: the carbon side of the grid was placed onto a drop of 2% uranyl acetate stain for 10 seconds, excess of liquid was blotted off.

Staining with NanoW™ (Nanoprobes): a drop of NanoW™ was placed onto the carbon side of the grid and incubated for 1 min, excess of liquid was blotted off. Alternatively, the drop was applied twice for 30 seconds each time.

Staining with PTA: the pH of PTA was adjusted with KOH to 7.4; the adjustment was done directly before the experiment as PTA is unstable at this pH. A drop of PTA pH 7.4 was placed onto the carbon grid and incubated for 1 min, excess of liquid was blotted off.

The grids were air-dried and mounted onto a Tecnai 10 or 12 FEI transmission electron microscope (accelerating voltage 120 kV) and analysed at a magnification of 42000 - 52000.

Chapter 3

Annotation of the NT-26 two-component "signalome"

Iterative BLASTp searches against the NT-26 genome using various protein and domain sequences identified 90 putative TCS proteins, 46 HKs and 44 RRs. Predicted domain types and arrangements were used for their classification and functional annotation. Where possible, proteins were assigned into putative TCSs, including the previously characterised AioSR system [112], and systems potentially unique to NT-26 were identified. A search for all other signal transduction proteins encoded in the NT-26 genome (e.g. methyl-accepting chemotaxis proteins) was conducted in order to compare the NT-26 "signalome" and signalling "intelligence" to those of other bacterial species, as well as to establish whether NT-26 senses mainly environmental or intracellular parameters.

Expression constructs were generated for a selected subset of 26 proteins, comprising 12 putatively assigned TCSs. These proteins were expressed and partially purified providing preliminary data for further characterisation of TCSs in NT-26.

3.1 TCS proteins - diversity of domain types and arrangements

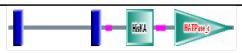
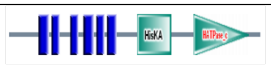
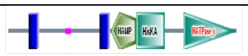
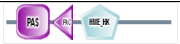
3.1.1 Histidine kinases

The list of all proteins containing a HK core is given in Table 3.2. Out of these, 24 were predicted to contain at least one transmembrane region while 22 were fully soluble. In total there were 12 hybrid HKs characterised by the presence of a receiver domain at the carboxyl end; two hybrids contained a tandem of receiver domains and one possessed

three receiver domains. Most HKs contained DHp domain of HisKA type (PF00512) followed by the catalytic domain (HATPase_c; PF02518). Two proteins were predicted to contain DHp domains of HisKA_3 type (PF07730) which were recognised by Pfam, but not by SMART. Eight proteins were predicted to possess a HWE_HK domain (PF07536), which in only three cases was followed by a HATPase_c domain - in five proteins assigned as HWE HKs no catalytic domain was recognised by SMART or Pfam.

The sequences of the predicted domains containing the phosphorylatable histidine residue (HisKA, HisKA_3 or HWE_HK) were used to assemble a multiple sequence alignment, which was examined to confirm the presence of the invariant histidine residue. The HWE_HK domain of *Agrobacterium tumefaciens* BphP2 (UniProt entry A9CI81) was included in the alignment to determine the position of the phosphorylatable histidine within HWE_HK domains; His523 was previously shown to constitute the phosphorylation site in BphP2 [71]. The sequences of NT-26 proteins containing HisKA_3 domains had to be aligned separately due to the presence of a different consensus sequence directly before the active site histidine. HisKA_3 domain of an uncharacterised *Bacillus cereus* HK (UniProt entry B3ZQM2), which is given as one of the representative HisKA_3 proteins by Pfam, was included in that alignment. All proteins identified as HKs in NT-26 contained the conserved histidine residue (see Appendix, Fig. B.1 and B.2) except for one cytoplasmic protein, with three PAS domains, which contained a tyrosine at that position. As tyrosine phosphorylation occurs in most organisms, it is conceivable that in HK3858 the tyrosine residue is phosphorylated. A multiple sequence alignment was also generated using all receiver domains detected in both hybrid HKs and RRs; each of the receiver domains associated with hybrid HKs was likely to be functional based on the presence of the phosphorylatable aspartate residue (see Appendix, Fig. B.3).

ID (name)	**Type**	**Domain arrangement**	**Length (a. a.)**	**RR ID (name)**	**Cognate RR type**
13	-		550	-	-
57	PER		498	-	-
79	CYTa		414	-	-
142	TM/C		1169	-	-

ID (name)	Type	Domain arrangement	Length (a. a.)	RR ID (name)	Cognate RR type
152	CYT		639	150	DNA
197 (CheA)	CheA		746	196, 200, 201	CheY, CheB
281	CYT		1192	-	-
656	PER		461	655	DNA
846	TM/C		507	-	-
917	PER		468	916	DNA
925	PER		470	924	DNA
927	TM/C		782	-	-
1631 (NtrB)	CYT		382	1632 (NtrC)	DNA
1633 (NtrY)	TM/C		756	1634 (NtrX)	DNA
1656	PER		384	1657	DNA
1689	TM/P		513	-	-
1761 (BphP2)	CYT		850	-	-
1898	CYTa		865	-	-
2202	CYT		968	-	-
2314	PER		470	2313	DNA
2464	CYT		399	-	-
2493	CYT		1223	-	-
2677	CYT		344	-	-
2679	PER		499	2678	CheY-like
2813	CYT		330	-	-
2858	-		342	2859	DNA

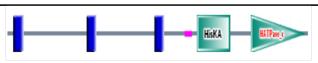
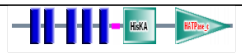
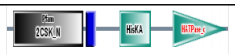
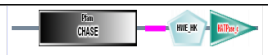
ID (name)	Type	Domain arrangement	Length (a. a.)	RR ID (name)	Cognate RR type
2892	CYT		638	-	-
2925	PER		569	-	-
2937	PER		614	2936	DNA
3039	PER		453	3038	DNA
3041	(PER)		455	3042	DNA
3223	TM/P		1125	3224	DNA
3236	CYT		413	3237	DNA
3252	CYT		681	-	-
3656	REC		577	-	-
3657	CYT		2102	3656	REC HK
3733 (BphP1)	CYT		732	(X)	(X)
3736 (ExsG)	REC		336	(X)	(X)
3760	PER		562	-	-
3858	CYT		814	-	-
3863 (ChvG)	PER		594	3864 (ChvI)	DNA
3885	TM/P		447	3884	CheY-like
4001	PER		462	4000	DNA
4221	CYT		501	-	-
4223	CYT		865	-	-
p10027 (AioS)	PER		489	p10028 (AioR)	DNA

Table 3.2: List of all NT-26 proteins containing a HK core. ID refers to Genoscope ID given to each gene in the sequencing project; for name assignment see Section 3.2. pXXX refers to genes encoded on NT-26 plasmid. Type classification was based on sensing: CYT - cytoplasmic, CYTa - cytoplasmic anchored, PER - periplasmic, TM/P - transmembrane/periplasmic, TM/C - transmembrane/cytoplasmic, REC - N-terminal receiver domain. Domain arrangements are represented using SMART graphical output; green lines denote predicted coiled-coil regions, pink - low complexity regions. Cognate RRs were identified based on gene co-localisation; types: DNA - DNA-binding domain, CheY/CheY-like - single-domain RR. (X) denotes AgR and ExsF which are encoded next to BphP1 and ExsG genes; the cognate pairs were identified in this project (see Chapter 4).

Among the cytoplasmic proteins one was annotated as the chemotaxis HK (CheA, ID 197) and two as bacteriophytochrome photoreceptors (BphPs; IDs 1761, 3733)). Three proteins displayed an atypical domain arrangement with a receiver domain appended at the N terminus ("REC"-type in Table 3.2), suggesting that their kinase cores may be regulated by phosphorylation of the preceding domain; this would imply that these HKs function as RRs (hybrid RRs, see Section 1.3.3). Unusually, two soluble HKs appeared to have no sensory regions (IDs 13, 2858), and one of them seemed to be a hybrid HK as a receiver domain was predicted at its C terminus. Another protein (ID 3657), comprising 2102 residues, contained multiple N-terminal HAMP domain (found in histidine kinases, adenylyl cyclases, methyl binding proteins and phosphatases; PF00672) and three C-terminal receiver domains, along with a GAF domain (found in cGMP-specific phosphodiesterases, adenylyl cyclases, FhlA; PF01590, PF13185 or PF13492) and a classical HK core. Despite the unusual domain composition and extraordinary size, this protein could constitute a functional HK as it contained four putative phosphorylatable residues (1 His and 3 Asp), and its homologues were identified in several other *Rhizobiaceae.*

The N-terminal sensory regions found in NT-26 HKs were classified based on their localisation (periplasmic, membrane-integral/periplasmic or cytoplasmic) as well as the number and types of domains involved; the frequencies of domains found in at least two different HKs are shown in Figure 3.1. Eight out of 24 proteins with transmembrane helices were predicted to contain a HAMP domain, which has been implicated in signal transmission from the N-terminal sensing region to the downstream HK core [27, 128, 4].

In most HKs stimulus detection seemed to be carried out by the periplasmic region of the protein. There were 14 proteins which were predicted to contain a periplasmic domain flanked by two transmembrane helices of which one leads into the HK core. Out of these, one appeared to have a CHASE-type domain (CHASE3, cyclases/histidine kinases associated sensory extracellular 3; PF05227) and one was predicted to contain

a 2CSK_N domain (PF08521). In others the extracytoplasmic region has not been recognised as a domain by SMART or Pfam databases. Curiously, one protein (HK2937) was predicted to contain a third transmembrane helix dividing the sensory region into two sections (∼130 and ∼120 amino acids); such arrangement suggests that the sensing region could comprise a periplasmic as well as a cytoplasmic domain.

Three HKs possessed multiple (6 - 14) transmembrane helices and no obvious periplasmic portions. Such domains could be involved in sensing membrane-related changes, but they could also detect periplasmic stimuli, even in the absence of a considerable periplasmic domain [19]; hence they were termed "transmembrane/periplasmic". No recognisable domains were detected within the sensory regions of these proteins.

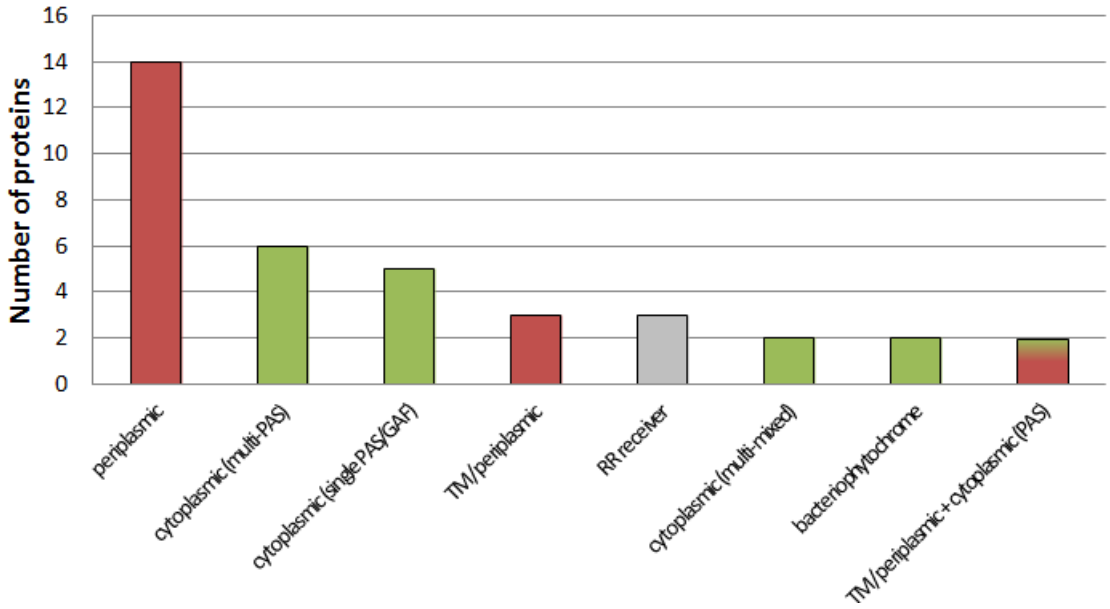


Figure 3.1: Most common domain arrangements within HK N-terminal portions. Green denotes regions involved in cytoplasmic sensing, red - in periplasmic or transmembrane sensing. RR receiver domains belong to soluble proteins but are shown in grey as these proteins may act as RRs, not HKs.

Among the 19 cytoplasmic sensor proteins the N-terminal region most commonly contained the PAS domain (period, aryl-hydrocarbon receptor nuclear translocator and single-minded; PAS - PAS_9 in Pfam), existing as a single domain (3 HKs) or as an array of two or more PAS domains (6 HKs). Notably, as the SMART database does not contain all HMM profiles for the 9 types of the PAS fold existing in the Pfam database, all HK sequences were also scanned against the Pfam database [65] to ensure that all PAS domains are recognised. In most cases SMART and Pfam predictions of PAS domains were in agreement with each other but, interestingly, there was one PAS domain that was detected by SMART (ID SM00091) and not by Pfam. There

were 2 cytoplasmic proteins with a single GAF domain, one with PAS + GAF domains and two BphPs with a characteristic PAS + GAF + PHY (phytochrome; PF00360) arrangement. In addition, a GAF domain was identified in the unusually long HK with multiple HAMP and receiver domains. Curiously, one soluble HWE HK (ID 4221) contained a CHASE domain (PF03924), which is normally located in the periplasm, suggesting that CHASE domains could be involved in sensing cytoplasmic stimuli. The remaining 3 proteins belonging to this class were the HKs with N-terminal receiver domains.

Six HKs encoded within NT-26 genome contained both cytoplasmic sensing domains and at least one transmembrane region. Two proteins were likely to be membrane-anchored cytoplasmic sensors (CYTa in Table 3.2); one contained a single transmembrane helix and two PAS domains while the other, a hybrid HK, had two consecutive transmembrane helices followed by three PAS domains. Membrane attachment may be required for correct positioning of the cytoplasmic sensory region, as in the case of the turgor-sensing HK KdpD [129]. The remaining four proteins were assigned as putative transmembrane/cytoplasmic sensors. One of them (ID 846) contained two N-terminal transmembrane helices, but the ∼150 residue region between them and the HK core was not recognised as a domain; the considerable length of this cytoplasmic portion preceding the HK core suggested that it may be involved in sensing. Three other proteins possessed membrane-spanning or periplasmic portions as well as one or two PAS domains on the cytoplasmic side; such domain arrangement suggests that these proteins may be involved in monitoring both extracellular and cytoplasmic parameters.

Finally, there was one protein (ID 3041) which was predicted to contain a 2CSK_N domain and only one transmembrane region (constituting a part of 2CSK_N domain). The sequence of this protein was submitted to TMHMM server version 2.0 [130] and from the resulting plot it transpired that there could be another transmembrane helix starting approximately 130 residues after the first helix, but the prediction was not significant (the calculated probability was less than 1). Although it is likely that the protein belongs to the periplasmic-sensing HKs, it was counted separately.

3.1.2 Response regulators

NT-26 RRs were assigned based on the presence of the receiver domain (REC in SMART or Response_reg in Pfam) and a phosphorylatable aspartate residue, which was determined from the multiple sequence alignment of all predicted NT-26 receiver domains (see Appendix, Fig. B.3). In total 46 putative RRs were assigned, but two of them, containing C-terminal DNA-binding regions, seemed to lack the invariant aspartate re-

sidue, and were therefore not classified as RRs. However, such receiver domains may be exploited in phosphorylation-independent signalling, as in the case of *Pseudomonas aeruginosa* protein AmiR whose N-terminal portion adopts a receiver-like fold but lacks essential phosphotransfer residues [131]. AmiR transcription antitermination activity is regulated by protein-protein interactions with the ligand-binding AmiC.

The number of identified RRs (44) was slightly lower than the total number of HKs; however, as mentioned in the previous section, three NT-26 HKs seemed likely to function as RRs. This possibility was tested on one of these proteins and the results arere described in Chapter 6. NT-26 RRs were classified on the basis of the C-terminal effector domain (Fig. 3.3). Two atypical proteins were found; one contained a sensor domain (PAS) N-terminal to the receiver domain while the other comprised a tandem of receiver domains. In the latter protein the first receiver domain was unlikely to be functional as multiple sequence alignment of all receiver domains indicated lack of a phosphorylatable aspartate (see Appendix, Fig. B.3, 1994rec1).

DNA-binding	AAA+ ATPase + DNA-binding	Enzymatic	ANTAR	None (single-domain RR)	Unusual
21	5	2	1	13	2

Table 3.3: Summary of NT-26 RR types classified based on the effector domain. ANTAR - RNA-binding domain involved in transcription antitermination.

Most (26) RRs contained a DNA-binding domain of helix-turn-helix (HTH) type which suggests that majority of responses triggered by TCSs in NT-26 involve regulation of gene expression. The DNA-binding region was normally appended at the C terminus but one protein contained an N-terminal domain of sigma70_r4_2 type (sigma70-like sigma factor, region 4; PF08281). Five RRs possessed an AAA+ domain (ATPases associated with various cellular activities, PF07726) between the receiver and DNA-binding domains which has been shown to multimerise; such RRs activate transcription from σ^{54} promoters [55]. Several types of HTH motifs were found in NT-26 RRs with Trans_reg_c (transcriptional regulatory; PF00486) domain being the most prevalent. Trans_reg_c domains belong to the family of "winged HTH" transcription factors, which contain two β-sheets as well as the HTH motif.

The second largest group of RRs comprised CheY-like stand-alone receiver domain proteins which lack any effector domain. Such RRs are frequently involved in regulating bacterial motility or in multistep phosphorelays. In NT-26 there were 13 CheY-like proteins; two were encoded within the chemotaxis operon, two in a bacteriophytochrome gene cluster and one was encoded next to a gene coding for a canonical HK. The remai-

ning single-domain RRs constituted "orphan" proteins. The presence of two chemotaxis RRs in NT-26 is analogous to that in *Sinorhizobium meliloti*, where the two CheYs were hypothesised to form different interactions with the flagellar motor, or to participate in different signalling chains [132]. In *Rhodobacter sphaeroides* all six CheY proteins that are encoded within the chemotaxis operon were shown to bind to FliM, suggesting that they either compete for FliM binding or that some act as phosphate sinks and do not affect flagellar motor rotation [133]. Interestingly, the two *S. meliloti* CheYs are likely to be functional even though they are significantly divergent in terms of sequence [132]. This also applies to NT-26 CheYs; based on sequence alignment it appears that CheY1 and CheY2 share the same level of sequence identity with each other (35%) as with *E. coli* CheY (UniProt entry P0AE67). Both NT-26 CheYs contain the phosphorylatable aspartate and most of the other conserved amino acid motifs (see Appendix, Fig. B.3, 0196 and 0201)). However, CheY1 has a valine in place of the tyrosine involved in Y-T coupling, while CheY2 has a tryptophan at that position, whose aromatic side chain may imitate the behaviour of tyrosine during RR activation. Lack of an aromatic residue at this crucial site in CheY1 implies that this protein could bind to FliM with a different affinity, or compete with CheY2 for FliM binding site. Alternatively, it may not affect flagellar motor rotation but instead act as a phosphate sink, whereby it would reduce the phosphorylation level of the other CheY protein.

Two RRs contained effector domains with enzymatic activity. One comprised a tandem of receiver domains followed by a diguanylate cyclase domain (named GGDEF after five conserved residues; PF00990); multiple sequence alignment of all NT-26 receiver domains showed that the second receiver domain is not functional as it lacks the phosphorylatable Asp residue (see Appendix, Fig. B.3, 1293). The other protein was assigned as CheB chemoreceptor methylesterase based on the predicted protein-glutamate methylesterase domain (CheB_methylest; PF01339).

Finally, one RR contained an RNA-binding ANTAR domain (PF03861) involved in the regulation of transcription antitermination.

3.1.3 Histidine phosphotransfer domain

BLAST and PSI-BLAST searches against NT-26 genome identified only one Hpt domain other than the one present in CheA protein, although, given the low sequence conservation of Hpt domains [69], additional Hpt domains could have been missed out in the searches. The identified domain constituted a single-domain protein; it was not encoded in the vicinity of any other TCS protein genes and therefore it was impossible to determine in which signalling pathway it may be participating. Notably, in

the context of of an organism whose genome encodes 12 hybrid HKs, the presence of a single Hpt domain is striking.

It appears that in all prokaryotic genomes Hpt domains are under-represented in comparison to hybrid HKs [134]. The reason for such disproportion seems to be the need for signalling convergence - one Hpt domain can receive the signal from several different hybrid HKs, thus mediating a common response to a range of stimuli. As an example, *Pseudomonas aeruginosa* HptB, a single-domain Hpt protein, constitutes the convergence point for four different signalling pathways, and via a single cognate RR it regulates the expression of genes associated with biofilm formation and motility [135]. Convergence of signalling pathways also seems to occur in Eukaryotes. In the complex cytokinin signalling pathway in the plant *Arabidopsis thaliana* each of the five single-domain Hpt proteins can receive the phosphate group from several hybrid HKs and transfer it onto a number of different RRs [136].

The fact that NT-26 genome encodes 12 putative hybrid HKs but only one Hpt domain suggests that signal transduction initiated by the hybrid HKs may converge at the Hpt domain, and the phosphate group may be transferred onto one or several RRs. This would imply that, as in *P. aeruginosa*, in NT-26 the detection of several different stimuli may be coupled with a single response, or several parallel responses as in cytokinin signalling.

3.2 Assignment into functional systems

The primary criterion used in assigning proteins into phosphotransfer systems was gene co-localisation; proteins encoded in an operon or next to each other are likely to form a signalling pathway. Thus 21 systems, in addition to AioSR, were identified (see Table 3.2). All these TCSs seemed to have homologues in at least some members of the *Rhizobiaceae* family (with over 30% sequence identity for each system member).

3.2.1 Annotation of the names and functions of NT-26 TCSs

For NtrYX, NtrBC, FeuQP, ActSR and the chemotaxis system names and functions were assigned based on domain arrangement, operon or gene cluster structure, and a high level of sequence identity with previously characterised proteins. Chemotaxis proteins containing the domains involved in phosphotransfer (CheA, CheB, CheY) are frequently encoded in an operon with other Che proteins, which facilitated their identification within the NT-26 genome. CheB methylesterase domain, stand-alone receiver

domain protein CheY and the five-domain architecture of CheA are also characteristic features of this system. Chemotaxis TCS regulates bacterial motility in response to attractants or repellents. Proteins involved in nitrogen regulation (NtrYX and NtrBC) were also encoded in one operon and shared a high level of sequence identity and a common domain arrangement with equivalent well-characterised systems in other organisms. Their sequences were over 50% identical with other rhizobial homologues, and both RRs contained an AAA+ domain. In addition, NtrB possessed a PAS domain and NtrY was a membrane-embedded HK. Another NT-26 TCS, assigned as ActSR, is commonly involved in response to acidic conditions, and in *S. medicae* it was shown to regulate the expression of several different genes [137, 138]. The sequences of both proteins in NT-26 were over 70% identical with their *S. medicae* homologues. The FeuQP system was identified based on homology to a *Rhizobium leguminosarum* TCS involved in high affinity iron uptake [139]. FeuQ, a canonical HK with a periplasmic domain, shared 69% sequence identity and the transcriptional regulator FeuP shared 87% identity with *R. leguminosarum* homologues.

One of the identified TCSs exhibited over 80% sequence identity to ChvGI from *Agrobacterium tumefaciens* and ExoS-ChvI in *Rhizobium meliloti*. However, these TCSs are involved in virulence regulation (*A. tumefaciens*) [140] and exopolysaccharide production necessary for symbiotic interactions with the plant host (*R. meliloti*) [141]. As NT-26 is unlikely to interact with plants, the function of this system in NT-26 is unclear.

Finally, based on sequence identity of over 50% to *A. tumefaciens* homologues, names and functions were annotated for five other proteins, including two red and far-red light sensors - bacteriophytochrome photoreceptors (BphP) 1 and 2. As in *A. tumefaciens* [71], NT-26 BphP2 appears to be an orphan protein, while BphP1 is encoded in a gene cluster with AgR, ExsG and ExsF. Furthermore, the domain arrangements of all five proteins and the operonic organisation of the BphP1 gene cluster were identical to that in *A. tumefaciens*. It was previously shown that in *A. tumefaciens* BphP1 phosphorylates AgR [71], which supports the idea that the four proteins encoded in BphP1 gene cluster constitute a signalling pathway. The remaining 14 NT-26 TCSs could not be assigned as no homologues with established functions were identified.

3.2.2 Identification of NT-26 TCS(s) that are not found in other *Rhizobiaceae*

Given that NT-26 occupies an environmental niche which would be expected to differ from that occupied by other members of the *Rhizobiaceae* family, it seemed likely that its genome would encode some TCSs that would not be homologous to any TCSs present in its close relatives. Such TCSs, unique to NT-26 in the context of *Rhizobiaceae* species, could have been acquired horizontally from other inhabitants of the gold mine or evolved through evolutionary processes associated with gene duplication and rearrangements.

Surprisingly, there was no TCS unique to NT-26, but there was one pair of proteins (HK3041 and RR3042) for which BLAST searches identified homologues in only one other family member - *Agrobacterium tumefaciens* strain F2. *A. tumefaciens* F2 has been shown to produce bioflocculants, i.e. compounds that induce flocculation (aggregation) of colloids [142]. HK3041 shared 53% and RR3042 shared 68% sequence identity with the respective homologues, encoded by genes 12503062 and 12503063. They also shared over 50% sequence identity with homologous sequences in *Nitratireductor aquibiodomus* (*Phyllobacteriaceae*), a nitrate-reducing bacterium [143], and in the manganese oxidiser *Aurantimonas manganoxydans* (*Aurantimonodaceae*) [144]. Both these aquatic species are members of families that, along with *Rhizobiaceae*, belong to the order *Rhizobiales*.

HK3041 was predicted by SMART to constitute a transmembrane HK with an N-terminal 2CSK_N domain, encompassing one transmembrane helix, and a C-terminal classical HK core. As described in Section 3.1.1, it is likely that there is another transmembrane helix within the sensory region, but the prediction by TMHMM 2.0 was not significant enough to confidently assign this protein as a periplasmic-sensing HK. RR3042 comprised a receiver domain and a C-terminal DNA-binding domain of the winged HTH type (Trans_reg_c). Therefore 3041/3042 may form a canonical TCS which regulates gene expression in response to an unknown stimulus.

The sparse distribution among *Rhizobiales* suggests that 3041/3042 is not native to this order of α-proteobacteria. In the case of NT-26 it may have been acquired from other microorganisms inhabiting the gold mine. Gene duplication is unlikely to explain its emergence - HK3041 shares only 26% overall sequence identity with HK4001, the other classical HK containing the CSK_N domain. Interestingly, HK3041 and RR3042 are encoded next to another putative canonical TCS (HK3039, RR3038), separated by a gene encoding a putative transporter protein. In contrast to the 3041/3042 system, HK3039 and RR3038 have homologues in other members of the *Rhizobiaceae* family.

All these five genes are encoded in the same direction but, given that 3041/3042 was most likely acquired by horizontal gene transfer, it seems unlikely that they form an operon.

As the closest homologues of 3041 and 3042 were identified in species exhibiting diverse metabolic strategies, and none of the homologues have been functionally characterised, it was impossible to predict the stimulus to which this TCS responds. It appears, however, that this stimulus is present in both aquatic and terrestrial environments.

3.3 NT-26 "intelligence" and "extrovertness" compared to other bacterial species

In order to calculate signalome-based IQ of NT-26, using the method outlined by Galperin and co-workers [81, 145], it was necessary to establish the number of all signalling proteins that can be encountered in bacteria: methyl-accepting chemotaxis receptor proteins (MCPs), diguanylate cyclase/phosphodiesterase proteins, phosphohydrolases with a conserved HD motif (HDs), adenylyl cyclases (ACs), protein serine/threonine/tyrosine kinases (STYKs) and protein serine/threonine/tyrosine phosphatases 2C (PP2Cs). The IQ can then be calculated using the formula $IQ = 50\,(lnN_{SP} - 2lnN_T) + 700$ [145], where N_{SP} is the total number of signalling proteins and N_T is the total number of proteins encoded in the genome. Signalome census of 555 bacterial and archaeal species is available from National Centre for Biotechnology Information website[1].

MCPs contain a highly conserved methyl-accepting (MA) domain (MCPsignal; PF00015). Keyword searches against NT-26 genome annotations were performed to identify several chemoreceptor proteins whose MA domain sequences were used as queries in subsequent BLAST searches; all sequences were submitted to the SMART database to confirm the presence of the MA domain. Altogether 16 chemoreceptors were found, of which two were cytoplasmic and one contained only an N-terminal MA domain. The same approach was used in the search for proteins with diguanylate cyclase (GGDEF; PF00990) and diguanylate phosphodiesterase domains (EAL; PF00563). Both GGDEF and EAL domains are named after conserved amino acid motifs. 12 proteins (2 cytoplasmic and 10 membrane-embedded) contained both these domains; 4 cytoplasmic and 2 transmembrane proteins had only single GGDEF domains and one cytoplasmic protein had just the EAL domain. In addition, one previously iden-

[1] http://www.ncbi.nlm.nih.gov/Complete_Genomes/SignalCensus.html

tified RR encoded in NT-26 genome contained a GGDEF effector domain (see Section 3.1.2), but it was not counted here as it is not a sensor protein.

Using keyword searches, one membrane-associated protein was identified with a C-terminal STYK domain (PF00069) and a PP2C domain (PF00481) at the N terminus. However, when these two separate domains were used as queries in BLAST searches no other proteins with STYK or PP2C domains were detected. The STYK domain of a putative *A. tumefaciens* kinase protein (UniProt entry Q7CUN1) was used as a query against the NT-26 genome but no STYK proteins were found other than the single STYK/PP2C protein. To identify proteins with the phosphohydrolase domain (HD; PF01966) and adenylyl cyclase (Adenylyl_cycl; PF01295) the respective domains of *Rhizobium etli* phosphohydrolase (UniProt entry Q2K5M9) and *A. tumefaciens* adenylyl cyclase protein (UniProt entry A9CHM3) were used as BLAST queries. However, no proteins with these domain types were detected.

Species	HK	RR	MCP	GGDEF	GGDEF +EAL	EAL	HD	AC	PP2C/ STYK	TOTAL	N_T	IQ
Rhizobium NT-26	46	44	16	6	12	1	-	-	1	126	4771	**95**
Agrobacterium tumefaciens	53	55	20	14	13	1	1	3	3	163	5520	**95**
Bradyrhizobium japonicum	92	90	35	12	23	4	3	37	6	302	8317	**83**
Rhizobium etli	55	64	27	14	21	-	1	22	4	209	5963	**98**
Rhizobium leguminosarum	61	65	28	16	22	-	3	29	5	229	7143	**84**
Sinorhizobium medicae	51	62	9	5	13	1	-	25	2	162	6213	**81**
Sinorhizobium meliloti	48	56	9	6	11	2	-	28	3	163	6218	**81**
Wolinella succinogenes	39	37	31	12	9	3	2	1	5	139	2043	**184**
Escherichia coli	30	32	5	12	7	10	-	1	2	99	4146	**97**
Neisseria gonorrhoeae	4	2	-	-	-	-	-	-	1	7	2002	**37**

Table 3.4: Comparison of NT-26 signalome composition and IQ to other rhizobia and selected bacterial species. N_T denotes the total number of proteins encoded in the genome.

The total number of signalling proteins in NT-26 was therefore 128 (assuming that all the identified proteins are functional), which is less than in several well-characterised species belonging to the order *Rhizobiales* (Table 3.4); for instance, *Bradyrhizobium japonicum* (family *Bradyrhizobiaceae*) encodes over twice as many signalling proteins. NT-26 seems to lack adenylyl cyclases, which in some of its close relatives are highly abundant, and there seems to be only one TCS which is unique to NT-26. The number

of MCPs in NT-26 is average compared to other *Rhizobiales* and to all other bacteria [146]. However, the number of proteins encoded by the NT-26 genome (4771) is also smaller than the protein complement of other rhizobia, and its signalling intelligence level (IQ = 95) is therefore the same as that of its close relative, *A. tumefaciens*. It appears that despite being mostly free-living organisms, members of the *Rhizobiales* order do not encode a considerable proportion of signalling proteins - the census of all sequenced bacterial genomes suggests that an IQ of 95 is below average. In fact, it is similar to that of symbiotic *E. coli* and two-fold lower than that of *Wolinella succinogenes*, the species with the highest signalling IQ. However, the IQ of NT-26 is still higher than that of several other bacteria, primarily pathogens (e.g. *Neisseria gonorrhoeae*) which do not require complex signal transduction in their relatively stable environments.

In order to assign NT-26 as an extrovert or introvert, the number of sensor proteins monitoring extracellular and cytoplasmic parameters was compared. To establish the total number of sensor proteins encoded by NT-26 genome, it was necessary to exclude those which do not have potential sensory regions at either N or C termini. In the context of TCSs this applied to CheA and two other HKs which lacked N-terminal sensory regions, as well as the three atypical hybrid HKs with N-terminal receiver domains which were assumed to act as RRs. One putative MCP, two GGDEF/EAL proteins and one EAL-only protein were also excluded as they contained just the signalling domains at the amino termini and no sensing domains were recognised elsewhere in their sequences. Hence the total number of NT-26 sensor proteins was 72. 10 GGDEF/EAL proteins, 2 GGDEF-only proteins, 13 MCPs and all transmembrane HKs, except for two membrane-anchored HKs likely to sense cytoplasmic stimuli (see Section 3.1.1), were counted as extracytoplasmic sensors. NtrB and both BphPs were also included in that category as they detect external stimuli. Therefore, although determination of the origin of each stimulus sensed by the sensor proteins would need experimental verification, it can be assumed that NT-26 genome encodes 50 proteins sensing external stimuli and 22 cytoplasmic sensors. As the majority of NT-26 sensor proteins are involved in monitoring external stimuli, NT-26 is a bacterial extrovert, monitoring primarily environmental changes. The emphasis on extracellular sensing is characteristic of organisms like NT-26, exhibiting several metabolic strategies [81, 82].

3.4 Cloning, expression and purification of selected TCS proteins

Out of 92 proteins that, based on their sequence similarity, were identified as possible components of signalling systems, 26 proteins belonging to 12 distinct systems were selected to carry out recombinant protein expression and purification trials. In the longer term these expression constructs could be used for characterisation of the NT-26 "signalome". One of these 12 TCSs was to be selected for further *in vitro* functional characterisation, provided all of its members could be expressed at a satisfactory level and purified.

3.4.1 Selection criteria for choosing the specific TC systems

In order to facilitate parallel cloning, the systems were selected based on the presence of restriction enzyme recognition sites within the genes. The vector used was pET30a which contains, among others, NcoI and HindIII restriction endonuclease sites, and these sites were to be exploited for insert ligation. NcoI and BspHI restriction endonucleases produce compatible "sticky" ends, hence BspHI could be used instead of NcoI to cleave those genes of interest that contained an internal NcoI site. If at least one gene encoding a member of the pathway contained HindIII and/or both BspHI and NcoI cleavage sites, the pathway was excluded. 12 pathways were thus chosen as listed in the table below (Table 3.5).

Several histidine kinases contained transmembrane helices and for these proteins only the portions of genes coding for the intracellular regions were included in generating expression constructs. The aim was to carry out parallel protein purification and inherently insoluble transmembrane proteins would require establishing individual purification protocols for each of the membrane proteins, thus adding significant complexity to the experiment. In addition, three response regulators (NtrC, NtrX and 2936) contained AAA+-type ATPase domains preceding the C-terminal DNA-binding domains; due to the propensity of AAA+ domains to oligomerise, which might have an effect on protein solubility and stability, only receiver domains of these RRs were cloned.

3.4.2 Cloning, expression and purification

All genes encoding selected proteins were successfully amplified from NT-26 genomic DNA using the protocol described in the Materials and methods, Section 2.2, and sub-

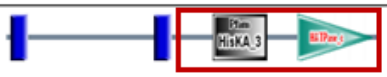
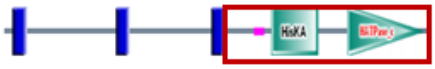
TCS	HK ID (name)	Domain arrangement	RR ID (name)	Domain arrangement
1	1631 (NtrB)		1632 $_{1-157}$ (NtrC)	
2	0925 $_{186-470}$		0924	
3	3236		3235	
4	1656 $_{109-384}$		1657	
5	0917 $_{241-468}$ (FeuQ)		0916 (FeuP)	
6	1633 $_{321-756}$ (NtrY)		1634 $_{1-130}$ (NtrX)	
7	3039 $_{211-453}$		3038	
8	3863 $_{301-594}$ (ChvG)		3864 (ChvI)	
9	4001 $_{191-462}$		4002	
10	2935 $_{321-614}$		2936 $_{1-140}$	
11	0152		0153	
12	3733 (BphP1)		3734 (AgR)	
	3735 (ExsG)		3736 (ExsF)	

Table 3.5: Domain arrangements and cloned portions of 26 TCS proteins selected for test expression. Gene ID refers to the ID given to each gene during the sequencing project, protein names are given wherever the putative function has been assigned. Domain arrangements are represented using SMART graphical output, including Pfam-annotated domains (in black). The second transmembrane helix in 1656, 0917, 3863 and 4001 is not indicated graphically due to partial overlap with the HAMP domain. Where only a part of the sequence was cloned, the portion is framed in red and amino acids incorporated in the construct are shown in subscript.

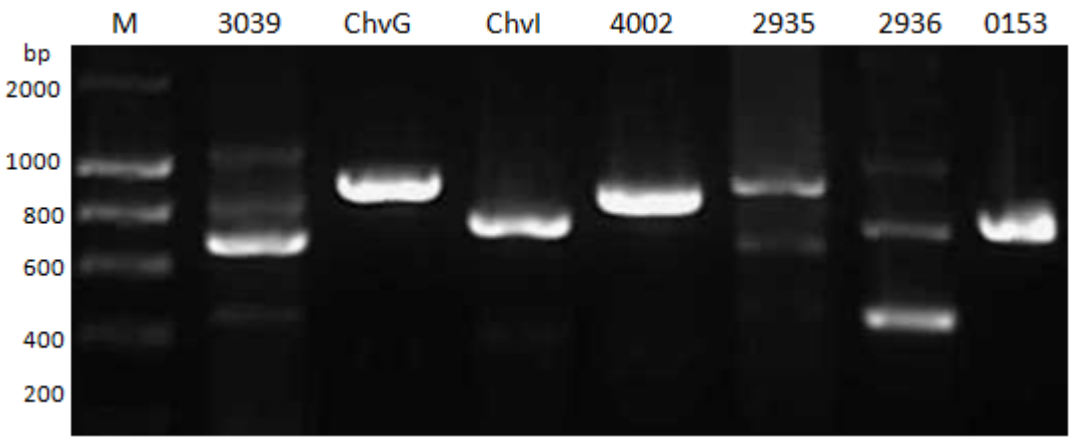


Figure 3.2: Example of PCR products. $3039_{211\text{-}453}$ - 672 bp, $ChvG_{301\text{-}594}$ - 885 bp, ChvI - 723 bp, 4002 - 678 bp, $2935_{321\text{-}614}$ - 885 bp, 2936 - 420 bp, 0153 - 735 bp. M - marker. 15 µl of each PCR reaction were loaded on the gel.

cloned into the expression vector. For the details of the multiple cloning site of pET30a plasmid see Figure A.1 in Appendix A. The resulting plasmids were purified and their sequence was verified. Figure 3.2 shows the results from several of PCR reactions showing the variability in purity and yields of PCR products obtained. Following PCR reactions, desired products were extracted from the gel, digested and ligated into pET30a vector downstream of T7 promoter sequence which is recognised by T7 RNA polymerase (T7Pol). Plasmid propagation was carried out in DH5α cells. After extracting plasmid DNA from DH5α cells and sequencing the region encoding the protein of interest, the expressing strain of *E. coli*, BL21 (DE3), was transformed. These cells possess T7Pol gene which is under the control of *lac* operon. During normal growth conditions the binding of the *lac* repressor to the *lac* promoter suppresses T7Pol expression, but upon induction with isopropyl-β-D-thiogalactopyranoside (IPTG) repressor binding is precluded and T7Pol is expressed. Each recombinant protein contained an approximately 4.5 kDa purification tag, incorporating a polyhistidine affinity tag, an S-protein binding site and a TEV protease cleavage site.

Small scale expression trials from 10 ml cultures showed that all proteins, except for BphP1 and AgR, were expressed and appeared soluble based on the SDS-PAGE following Ni^{2+} agarose-affinity purification protocol (Fig. 3.3). Each lane of the SDS-PAGE gels shown was loaded with 15 µl of the eluted fraction and a protein band should have been visible if the protein expression levels were at least 0.7 µg per ml of cell culture. Relatively poor levels of AgR expression were most likely caused by non-optimal choice of domain boundaries. At the time of this experiment, the NT-26 genomic sequence has not yet been finalised and subsequent data refinement showed

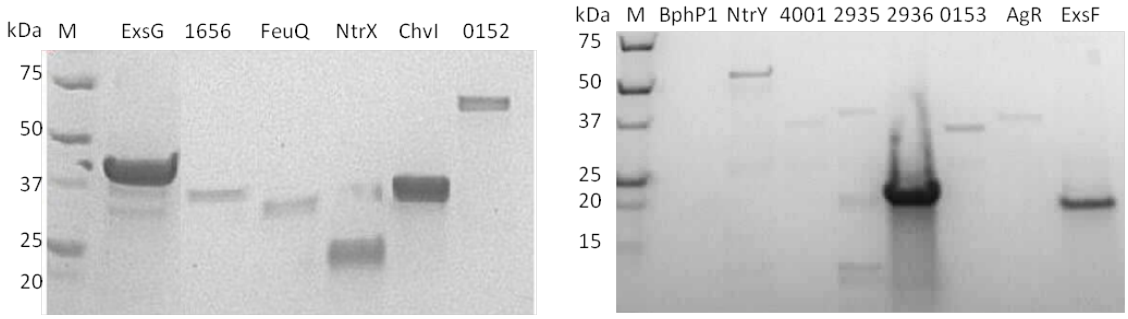


Figure 3.3: Example of test expressions. 12% SDS-PAGE gel. Each lane contained 15 µl of elution fractions of tagged proteins from NiNTA beads: ExsG (42 kDa), $1656_{109\text{-}384}$ (34), $NtrX_{1\text{-}130}$ (19), ChvI (32), 0152 (75), BphP1 (80), $NtrY_{321\text{-}756}$ (53), $4001_{191\text{-}462}$ (33), $2935_{321\text{-}614}$ (37), $2936_{1\text{-}140}$ (21), 0153 (34), AgR (21) and ExsF (20). M - marker.

that the AgR protein has a longer N-terminal sequence. This protein was later re-cloned, using the primers listed in the Materials and methods, Section (Table 2.1), and successfully expressed and purified (see Chapter 4).

In the case of BphP1, the protein remained in the insoluble fraction (Fig. 3.4). Notably, the proteins were expressed at 37°C which in the case of some proteins might have caused misfolding or unfolding thus leading to low yields.

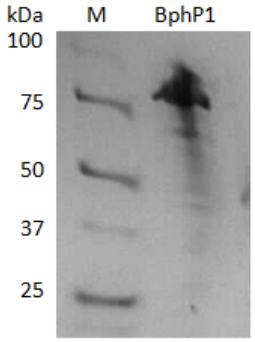


Figure 3.4: Insoluble fraction from BphP1 test purification. 12% SDS-PAGE gel. The lane contained 15 µl of insoluble fraction resuspended in the lysis buffer. BphP1 - 80 kDa.

3.5 Summary

The genome of *Rhizobium* NT-26 encodes nearly 90 putative two-component signalling proteins, some of which exhibit atypical domain composition and arrangement. The majority of HKs are associated with the membrane, suggesting that they most likely sense external stimuli, while most RRs seem to regulate gene expression. Nine out of 46 HKs were predicted to contain HWE or HisKA_3 variants of the HK core. Curiously,

12 hybrid HKs and only one single-domain Hpt protein were identified, which implies that several signal transduction pathways may converge upon the Hpt domain, thus evoking one type of response(s) to diverse stimuli.

Despite the fact that NT-26 occupies an environmental niche which differs from that occupied by several *Rhizobiaceae*, such as *A. tumefaciens*, its two-component "signalome" has not been significantly modified or enriched during evolutionary divergence. Only one TCS was identified as unique to NT-26 in the context of the *Rhizobiaceae* family; however, it was impossible to ascertain the stimulus it senses, as homologous proteins were found primarily in aquatic bacteria and have not been characterised. In the context of the entire "signalome" it appears that the signalling "intelligence" of NT-26 is average but comparable to that of its close relatives. As expected for an organism with several metabolic strategies, NT-26 is an "extrovert", with most sensor proteins involved in sensing extracellular stimuli.

22 TCSs were putatively assigned based on gene co-localisation. Out of these 8 were partially or fully annotated, including the putative phosphorelay which involves bacteriophytochrome photoreceptor 1. All of the 26 proteins selected for small-scale expression trials were expressed as soluble proteins, except BphP1 construct which was retained in the insoluble fraction. However, the apparent complexity of the putative signalling cascade initiated by BphP1, as well as the atypical domain arrangement and composition of ExsG, made this putative TCS an interesting target for further characterisation.

Chapter 4

Functional characterisation of the signalling pathway downstream of bacteriophytochrome photoreceptor 1

Results of the small scale expression trials presented in Chapter 3 indicated several proteins that expressed at a relatively high level. Among those, two proteins - ExsG and ExsF - were part of the BphP1 gene cluster. The other two proteins of the cluster, BphP1 and AgR, however, were retained mainly in the insoluble fraction. Subsequently, after an updated version of NT-26 genome was released, AgR was re-cloned and successfully purified. The kinase core portion of BphP1, sub-cloned and purified from soluble cell extracts (see below).

Given that the phosphorylatable domains of all its members were expressed as soluble proteins, and taking into consideration the unexplored signalling complexity of the BphP1-initiated pathway, this protein system appeared to be an appealing candidate for further characterisation. It offered the possibility of elucidating differences between canonical (BphP1) and HWE (ExsG) histidine kinases, understanding molecular recognition principles, and studying intricate signalling mechanisms. In addition, the unusual domain arrangement of ExsG, a hybrid RR with the N-terminal regulatory domain and the C-terminal kinase core, prompted a question regarding its potential role as it could function simultaneously as a HK and a RR. Finally, the response(s) triggered by BphP1 and BphP2 have not been yet elucidated due to the lack of associated effector domains; an attempt was therefore made to identify which cellular mechanisms may be regulated by bacteriophytochrome proteins.

4.1 Domain annotation and construct design

4.1.1 BphP1

The N-terminal sensory region of NT-26 BphP1 was predicted by SMART to contain three separate domains (Fig. 4.1A): PAS (period, aryl-hydrocarbon receptor nuclear translocator and single-minded; residues 11-107), GAF (cGMP-specific phosphodiesterases, adenylyl cyclase, FhlA; 131-309) and PHY (phytochrome, 310-490). A canonical form of HK core, comprising HisKA and HATPase domains, was assigned at the carboxyl end. The protein shared 55% overall sequence identity with *A. tumefaciens* BphP1 (UniProt entry Q7CY45).

4.1.1.1 Sensory region

The arrangement of the three domains was analogous to that in *A. tumefaciens* BphP1 [71]. Bacteriophytochromes contain two residues which serve as attachment sites for the chromophore: a cysteine near the amino terminus and a histidine residue within the GAF domain. These attachment sites were identified as Cys11 and His239 within NT-26 BphP1 sequence based on sequence alignment with its *A. tumefaciens* homologue. Therefore NT-26 BphP1 was likely to act as a red and far-red light sensor, in analogy to the characterised BphP1 from *A. tumefaciens*.

4.1.1.2 Histidine kinase core

BphP1 kinase core exhibited all the invariant sequence motifs characteristic of canonical HKs and these motifs were mostly identical with those present in the *A. tumefaciens* homologue (Fig. 4.1B). The putative phosphorylatable histidine (His515) was located within the DHp domain, while residues defining the ATP-binding cleft could be found within the CA domain.

4.1.1.3 BphP1-HK protein construct

As described in the Chapter 3, full-length BphP1 protein (732 residues) was retained in the insoluble fraction, while *A. tumefaciens* BphP1 has previously been successfully expressed as a full-length soluble protein. This could be due to the temperature at which the protein was expressed, or the composition of buffers used in purification. However, as the focus in this project was on the domains involved in phosphotransfer, purification of the full-length protein was not attempted again. Instead, the kinase core portion of BphP1 (termed BphP1-HK, residues 491-732) was cloned and purified separately (Fig.

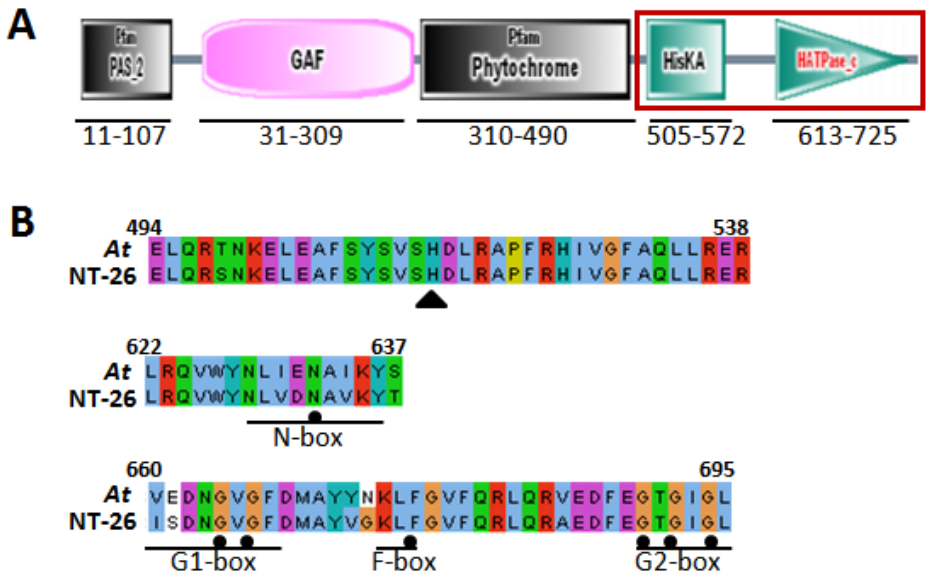


Figure 4.1: Domain arrangement and conserved residue motifs within NT-26 BphP1. A - BphP1 domain architecture predicted by SMART. The kinase core, which was sub-cloned separately as BphP1-HK construct, is framed in red. B - partial sequence alignment of *A. tumefaciens* (*At*) and NT-26 BphP1 kinase cores showing the region containing phosphorylatable histidine (black arrowhead), N-box, and the remaining three boxes (sequence motifs). The highly conserved residues from which the name of each box is derived are indicated by black dots. The sequences are numbered according to *At*BphP1 residues and coloured using ClustalX scheme.

4.1A). This expression construct incorporated a part of the linker extending from the PHY domain. An additional advantage of using the kinase core on its own was that the potential effect of light irradiation on autokinase activity was avoided.

4.1.2 ExsG

ExsG is an atypical type of histidine kinase, both because of its HWE HK core and the presence of an N-terminal receiver domain. Commonly, when a receiver domain is encoded within the same gene with a HK core, it is located at the carboxyl end of the protein. The reverse arrangement observed in ExsG and lack of a sensory region suggests that the activity of the catalytic core could be regulated by the receiver domain. NT-26 ExsG shares 56% overall sequence identity with its homologue from *A. tumefaciens* (UniProt entry Q7CY47) whose HK activity has previously been demonstrated [60].

4.1.2.1 HWE histidine kinase core

Classical HKs possess a DHp domain, also known as HisKA in Pfam and SMART, and a highly conserved C-terminal catalytic and ATP-binding region (CA domain) referred to as HATPase_c domain by SMART and Pfam. The analysis of the kinase core of ExsG by SMART identified a histidine-containing domain of the HWE type (HWE_HK in SMART) followed by a HATPase_c region. However, annotation of the latter domain was based on a rather high E-value (8.99), implying only distant homology of ExsG to HATPase_c domains associated with canonical HKs. In fact, in the case of 4 other HWE HKs found in NT-26, including bacteriophytochrome photoreceptor 2 (BphP2, 35% sequence identity with ExsG in terms of the kinase core), SMART does not detect any HATPase_c region C-terminal of the assigned HWE-HK domain. Interestingly, SMART also does not identify a HATPase_c region in *A. tumefaciens* ExsG. It appears that *A. tumefaciens* ExsG aligns both with HWE HKs and group 11 classical HKs [71]; the two kinase families share some level of similarity in terms of their glycine-rich (G-)boxes which may explain why the region C-terminal to the NT-26 ExsG HWE domain was annotated as HATPase, even though with a low confidence score. In addition, NT-26 ExsG does not contain the histidine residue that along with the WxE motif was used by Karniol and Vierstra to define the HWE family [71], while its *A. tumefaciens* homologue as well as NT-26 BphP2 possess all three residues. The absence of the specific HWE-motif histidine residue raised the question whether NT-26 ExsG is an active kinase; mutation of an equivalent residue (His616) in *A. tumefaciens* BphP2 (UniProt entry A9CI81) rendered the protein inactive. For clarity, the HATPase region of ExsG will be referred to as the "ATP-binding domain" (ABD), to account for the low level of ExsG HATPase similarity to typical catalytic domains (Fig. 4.2).

Alignment of the kinase core portions of NT-26 and *A. tumefaciens* ExsG and NT-26 BphP1 (Fig. 4.2B) showed that the histidine-containing region of ExsG HWE_HK domain shares some level of similarity with that of HisKA domains, but the putative phosphorylatable histidine in both ExsG sequences is followed by an arginine rather than an aspartate or glutamate. The presence of a highly conserved acidic amino acid at that position was shown to be crucial for autokinase activity across all HKs [147]. The N- and G2-boxes also share some similarity, except for one glycine of the consensus G2-box GxGxG motif being replaced with an acidic amino acid in ExsG. However, within the G1-box only the aspartate residue seems to be invariant across these three sequences, while the F-box seems to be absent from both ExsG proteins. In addition, two five-residue gaps in ExsG sequences aligned with BphP1 kinase core suggested that the spatial arrangement of the nucleotide-binding residues could be

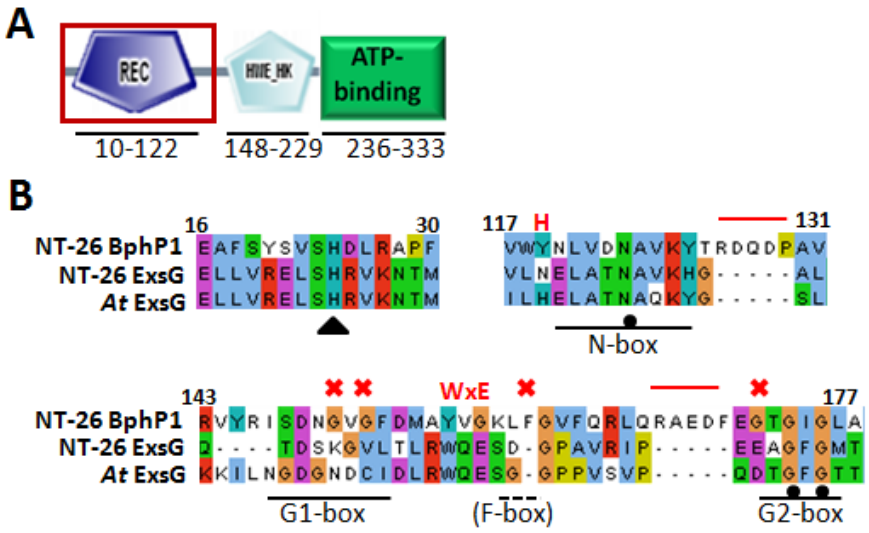


Figure 4.2: Domain arrangement of NT-26 ExsG and comparison of ExsG HWE_HK core to BphP1 HisKA core. A - ExsG domain architecture as predicted by SMART; the C-terminal HATPase region associated with an insignificant E-value was replaced with an "ATP-binding" domain. The receiver domain, sub-cloned separately to generate ExsG-REC expression construct, is framed in red. B - amino acid sequence alignment of NT-26 ExsG, *A. tumefaciens* (*At*) ExsG and NT-26 BphP1 kinase core (ClustalX colouring scheme). The position of the phosphorylatable histidine (black arrowhead) and conserved sequence motifs within the ATP-binding region are based on and numbered after the canonical kinase core (BphP1). Conserved residues present in both types of kinase core are indicated by black dots and those missing from HWE cores are indicated with red crosses. The characteristic HWE residues are in red above the sequences and red lines denote the characteristic gaps within ExsG sequences.

different from the one observed in canonical HATPase_c domains, and the ATP lid, normally extending between F- and G2 boxes, may be significantly smaller in ExsG proteins, if not completely absent.

4.1.2.2 Receiver domain

Aside from the phosphorylatable aspartate, several conserved residues have been identified in receiver domains [43] and ExsG N-terminal receiver domain sequence was examined in order to confirm their presence. As evident from the sequence alignment with the canonical single-domain RR, *E. coli* CheY (UniProt entry P0AE67), as well as AgR and ExsF (Fig 4.3), Asp62 constituted the putative phosphorylatable aspartate, as it is positioned directly after β3. Almost all highly conserved residues are found within ExsG receiver region, except for the proline of the KP motif that normally follows the fifth β-strand. In addition, based on the PSIPRED secondary structure prediction [118], it appears that strand β5 is absent - a coiled-coil region is predicted instead.

Considering that the domain most likely interacts with the appended HWE HK core, and β5 frequently interacts with the C-terminal effector domains of canonical response regulators, lack of this strand may reflect different type of interaction between the two functional domains. It has been reported that interdomain contacts mediated by the α4/β5/α5 interface often contribute to stabilisation of inactive RR conformations [75], but nothing is known about this relationship in such unusual hybrid HKs as ExsG. However, the fact that PSIPRED does not predict β5 does not exclude the possibility that this strand is present in ExsG structure - in fact, another secondary structure prediction programme, Jpred3 [119], assigned a β-type Ramachandran value to a single amino acid within that region.

4.1.2.3 WT ExsG and ExsG-REC protein constructs

Some of the experiments discussed in this Chapter involved two ExsG protein expression constructs - wild-type (WT) ExsG and its receiver domain purified as a separate protein. The receiver domain portion of ExsG was cloned, expressed and purified separately to establish if the receiver domain on its own is capable of receiving the phosphoryl group. This construct, termed ExsG-REC (residues 1-129), included the entire receiver domain and a part of the linker between receiver and HWE domains, as shown in Figure 4.2A.

4.1.3 AgR and ExsF

Both RRs encoded within the BphP1 gene cluster are single-domain RRs, comprising just the receiver domains. Comparison of AgR and ExsF sequences to ExsG-REC and *E. coli* CheY, revealed that despite overall similarity the two RRs exhibit certain unusual features (Fig. 4.3). Secondary structure prediction by PSIPRED indicated that all these proteins display the alternating β/α fold, with the exception that ExsG receiver domain lacks the fifth β-strand, as mentioned in Section 4.1.2.2. However, AgR appeared to have a longer loop region connecting α2 and β3 and an extended C-terminal portion compared to the other proteins, while ExsF lacked two amino acids from the set of the highly conserved residues. The structural details and putative functions of these two RRs will be discussed in more detail in Section 4.5.2. *E. coli* CheY is phosphorylated by its cognate HK (CheA) on Asp57, which aligns with ExsG-REC Asp62, ExsF Asp52 and AgR Asp66 - all these residues were located immediately after the third β-strand, which suggested that they constitute phosphorylation sites in the respective proteins.

Figure 4.3: Comparison of ExsF and AgR to ExsG receiver domain and *E. coli* CheY in terms of secondary structure elements and conserved residues. The sequences were aligned with respect to secondary structures and conserved amino acid motifs. CheY secondary structure elements were based on the crystal structure (PDB 3CHY); ExsF, AgR and ExsG receiver domain secondary structure elements were predicted by PSIPRED. Strands are shown in blue, helices in red, phosphorylatable aspartate residue is highlighted in yellow, highly conserved residues characteristic of receiver domains - in grey. ExsF residues which were different from the conserved amino acids at that position are highlighted in pink. AgR sequence extensions are framed in black.

4.2 Overexpression and purification of ExsG, ExsG-REC, BphP1-HK, AgR and ExsF

4.2.1 Test of BphP1-HK, AgR and ExsG-REC expression

To test the expression of the new constructs - BphP1-HK, ExsG-REC and AgR (re-cloned after final NT-26 genome data release), 2 litre LB cultures of BL21 (DE3) cells expressing each construct were grown at 37°C until the optical density measured at 600 nm reached 0.5 - 0.8. Protein expression was then induced by adding IPTG and the cells were incubated at 23 - 25°C for 4 hours. 15 µl of the cell culture prior to induction and 1, 2 and 4 hours after induction, was taken out and loaded onto an SDS-PAGE gel. In the case of each construct a protein band of corresponding size (including the purification tag - ∼5 kDa) was observed after induction (Fig 4.4).

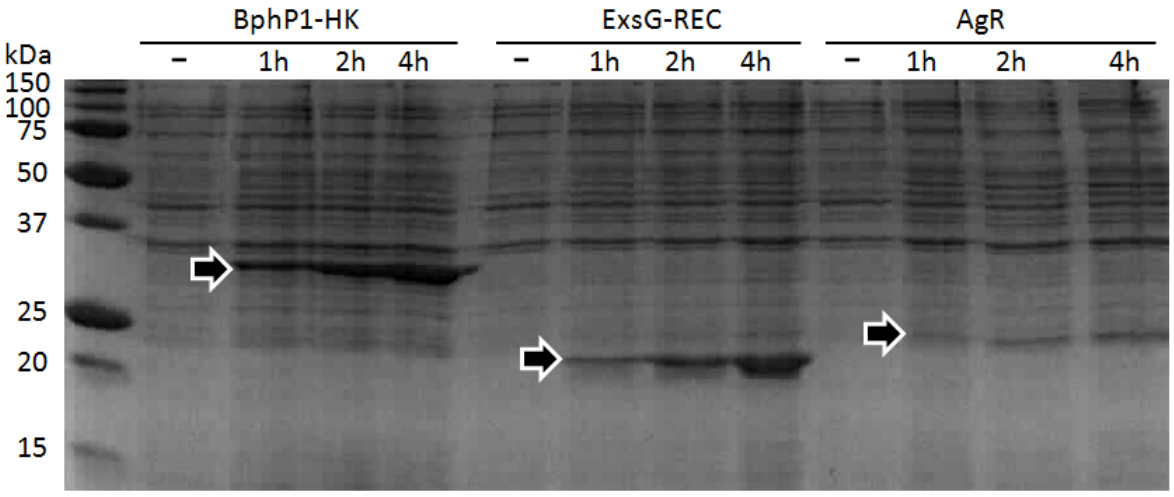


Figure 4.4: SDS-PAGE analysis of BphP1-HK, AgR and ExsG-REC protein overexpression. The arrows indicate protein bands appearing after induction of protein expression. BphP1-HK - 32.3 kDa, AgR - 22.3 kDa, ExsG-REC - 19.8 kDa.

4.2.2 Large-scale expression and NiNTA purification

Wild-type (WT) ExsG, ExsG-REC, BphP1-HK, AgR and ExsF were all expressed using the pET expression system which was described in Chapter 3. Proteins purified using Ni^{2+} agarose-affinity protocol were subsequently digested with TEV protease, after which two residues of the tag (Gly and Ala) would be retained at the N terminus of the protein. The required amount of NiNTA beads and TEV protease was established for each protein based on the preliminary purification results.

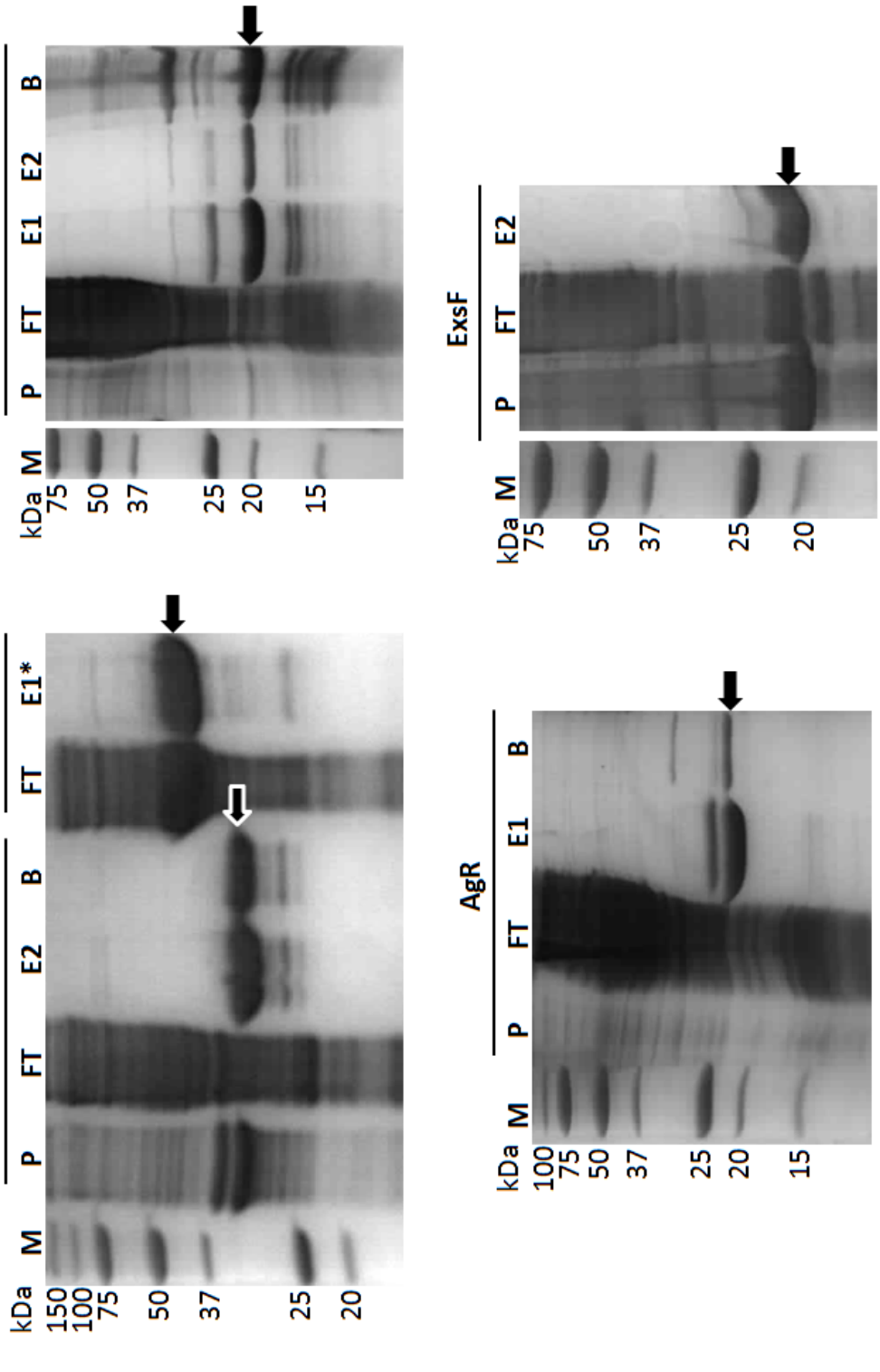


Figure 4.5: SDS-PAGE analyses of NiNTA purification of BphP1-HK, WT ExsG, ExsG-REC, AgR and ExsF. Arrows indicate the appropriate protein bands: BphP1-HK (31.5 kDa), WT ExsG (42 kDa), ExsG-REC (19 kDa), AgR (21.5 kDa), ExsF (18.5 kDa). M - marker, P - pellet (insoluble fraction), FT - flow-through, E - elution, B - beads fraction. WT ExsG elution, marked with asterisk, was diluted 4-fold prior to loading onto the gel.

Large-scale expression of WT ExsG, ExsG-REC, BphP1-HK, AgR and ExsF involved growing 2 - 4 litre LB cultures of BL21 (DE3) cells per construct until the optical density measured at 600 nm reached 0.6 - 0.8. Protein expression was then induced with IPTG and the cells were incubated for a few hours at 30 °C (WT ExsG) or 23 - 25°C (all other constructs). Cells were harvested, lysed and purified as described in Materials and methods (Section 2.4). Figure 4.5 shows sample SDS-PAGE analyses of NiNTA purification of each protein. It appeared that a large proportion of BphP1-HK and ExsF was retained within the insoluble fraction, but the amount of soluble protein obtained was more than sufficient for the planned experiments. WT ExsG purification is discussed in more detail in Chapter 5, but, as shown in Figure 4.5, the expression level of this protein was comparatively high. 2 - 4 ml of NiNTA beads were typically added to the lysate obtained from 12 g of cells, and the amount of protein eluted from the beads was between 50 and 120 mg, while over 200 mg still remained in the flow-through fraction.

Protein construct	Number of residues	Molar mass [kDa]	Theoretical pI	Theoretical extinction coefficient at 280 nm [M^{-1} cm^{-1}]	Average yield per 1 g of cells [mg]
$BphP1_{462-703}$ **(BphP1-HK)**	244	27	6.3	25440	4
WT ExsG	338	37.65	5.2	26930	20
$ExsG_{1-129}$ **(ExsG-REC)**	131	14.65	4.6	5960	1
AgR	154	17.16	4.8	9970	0.8
ExsF	127	13.8	4.5	8480	1.2

Table 4.1: Size, molar mass, isoelectric point, extinction coefficient and average purification yield of each protein product. The parameters were obtained using the online ProtParam tool available from ExPASy Bioinformatics Resource Portal [148]. The average yield was based on at least three rounds of protein purification.

Table 4.1 summarises the properties of each protein construct and provides the average yield from NiNTA purification. Notably, a larger volume of NiNTA beads (1 - 2 ml for protein obtained from 2 litre culture), and a larger amount of TEV protease (1 mg protease per 4 mg protein), was required for ExsF purification. It seemed that in the ExsF fusion protein the purification tag was occluded, limiting the access to the TEV recognition sequence and interfering with the affinity of poly-histidine tag for nickel ions. After removal of the purification tag by digestion with TEV protease, the

proteins were examined using SDS-PAGE to assess the purity of the proteins following which they were concentrated and subjected to size-exclusion chromatography.

4.2.3 Size-exclusion chromatography profiles

All proteins were soluble with and without the purification tag; however, after tag removal BphP1-HK appeared to be less soluble and more prone to precipitation than the tagged form. As soon as its concentration exceeded 4 mg/ml, heavy precipitation occurred, and the process could not be reversed by dilution. In contrast, the tagged form of BphP1-HK could be concentrated at above 5 mg/ml without apparent loss of protein due to precipitation. In addition, SEC profiles indicated that BphP1-HK was prone to forming high molecular weight aggregates (red arrow in Fig. 4.6), which were not observed in the case of the tagged protein - although both proteins seemed to adopt a range of oligomeric states (indicated by green and blue arrows in Fig. 4.6). Only the fractions that corresponded to the main absorbance peak were pooled.

WT ExsG elution profile consistently showed one uniform absorbance peak, suggesting that the protein exists in one native state; however, SDS-PAGE analysis of the protein after this purification step revealed the presence of two additional protein bands, which were of similar molecular weight to ExsG (Fig. 4.6). In order to separate those bands from the main protein band, electrophoresis time was extended by additional 20 min after the loading dye front reached the bottom of the gel. The proteins contained within the two bands were later shown to represent covalently modified forms of ExsG (see Chapter 5), and their abundance with respect to the unmodified protein differred in a prep-dependent manner. In addition, after removal of the purification tag, WT ExsG appeared to migrate further in SDS-PAGE gels than would be expected for a 37.6 kDa protein - the apparent mass was consistently lower by a few kDa. Determination of the monomeric and native mass of ExsG is described in Chapter 5.

The SEC profiles of the three stand-alone receiver domain proteins (AgR, ExsF and ExsG-REC) are shown in Figure 4.7. It seemed that both AgR and ExsF are likely to adopt a single conformation, indicated by the presence of one absorbance peak. In the case of AgR, SDS-PAGE analysis of the eluted protein indicated the presence of a contaminant (molecular weight 25-30 kDa) which could not be removed by the purification methods applied here. ExsG-REC seemed to exist in an equilibrium between two oligomeric states as its elution profile consistently comprised two peaks. SDS-PAGE analysis of protein samples taken from the relevant fractions showed that the protein corresponding to the first absorbance peak was partly degraded; hence only the fractions corresponding to the second peak were pooled. Notably, the absorbance

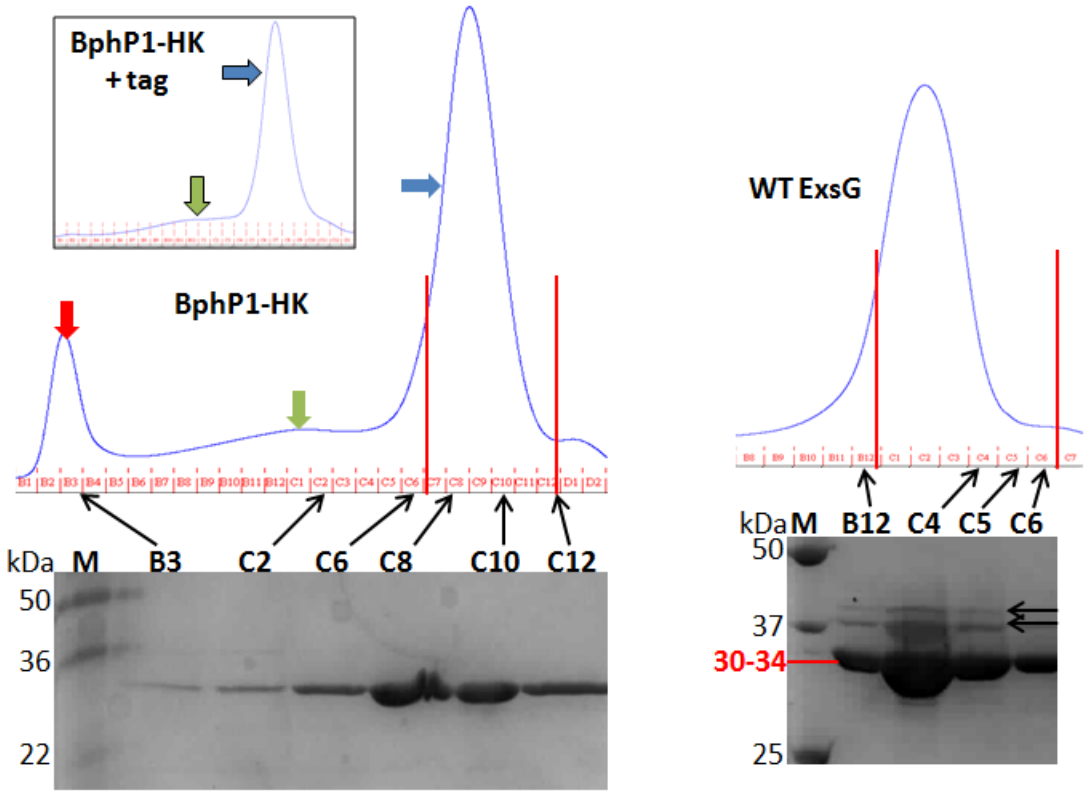


Figure 4.6: Purification of BphP1-HK and ExsG by size-exclusion chromatography. Vertical red lines on the partial elution profiles denote the range of pooled fractions. Partial BphP1-HK SEC profile (top) shows the main protein peak (blue arrow), a peak corresponding to aggregated protein eluted in void volume (red arrow) and protein possibly adopting a different conformation (green arrow). Partial SEC profile of BphP1-HK with the purification tag is shown in the inset. SDS-PAGE analysis of the each protein contained in the respective fractions is shown below. The apparent molecular weight of ExsG is shown in red, black arrows point at the contaminating bands. M - marker. BphP1-HK molecular mass - 27 kDa, ExsG (expected) - 37.6 kDa.

peaks of AgR and ExsF, as well as the larger ExsG-REC peak in the elution profile, correlated with a similar elution volume (18.5 - 19 ml); given that the molecular weights of these proteins are similar, it seemed likely that they adopt the same oligomeric state, although a proportion of ExsG-REC molecules appeared to form a higher order structure.

The molecular weights of all the above proteins were not determined using SEC but, despite the low resolution of the Superose 6 column, it was apparent that ExsG eluted earlier (14 ml) than BphP1-HK (16.5 - 17 ml) and the three receiver domain proteins (18.5 - 20 ml). Furthermore, ExsG elution occurred earlier than would be expected for a monomer or a dimer - the molecular weight based on the elution profiles of molecular weight markers was estimated to be over 400 kDa (see Chapter 5), suggesting that

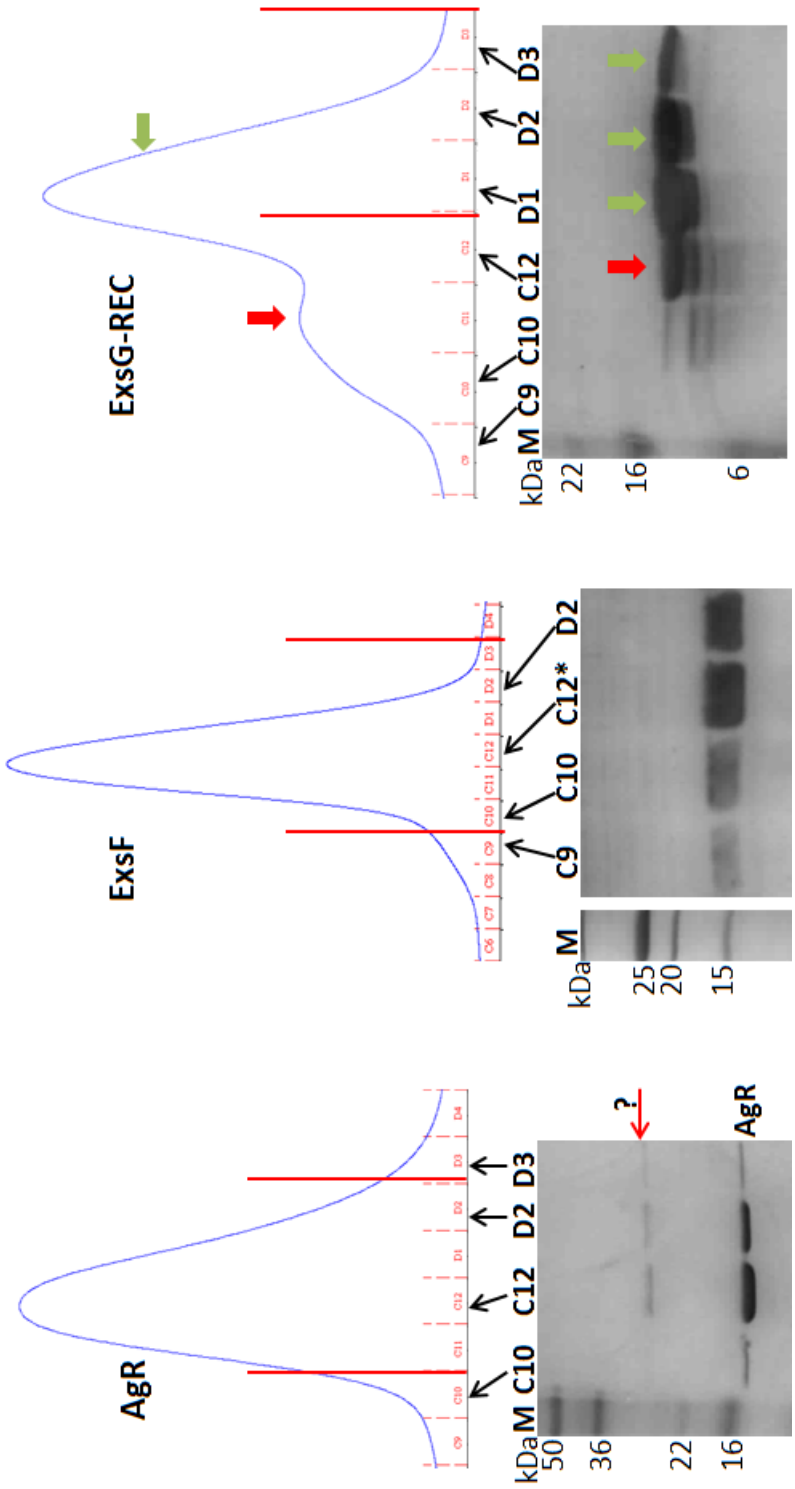


Figure 4.7: Purification of AgR, ExsF and ExsG-REC by size-exclusion chromatography. Red vertical lines on the partial SEC profiles represent the ranges of pooled fractions. Thin red arrow indicates the position of the AgR-associated contaminant. Thick red and green arrows indicate which gel lanes correspond to which chromatogram peaks. AgR - 17 kDa, ExsF - 13.8 kDa, ExsG-REC - 14.7 kDa.

this protein forms an oligomeric complex. Other experimental techniques were applied to determine the molecular mass and assembly pathway of ExsG, and to estimate the molecular masses of the other proteins, as described in Chapter 5.

4.3 Characterisation of BphP1-HK and WT ExsG autokinase activity and phosphohistidine stability

The two recombinant proteins, BphP1-HK and WT ExsG, were assayed for the ability to autophosphorylate over time and in the presence of varying ATP concentrations. Temporal and acid/base stability of the phosphoresidues were examined to confirm the involvement of a histidine residue, and to compare the rate of phosphate hydrolysis by canonical and HWE kinase cores. Furthermore, the active site histidine of WT ExsG was identified by testing autokinase activity of two ExsG histidine mutants.

All assays employed radioactively labelled ATP which was added as a tracer to the reaction buffer, as described in Materials and methods (Section 2.6.2); the buffer is herein referred to as the "hot" reaction buffer. Proteins were then subjected to SDS-PAGE electrophoresis and the localisation of radioactive signal on the SDS-PAGE gel was determined using phosphorimaging. The gel was placed underneath a phosphor screen, which "stores" the energy of the radioactive signal within the protein bands. The autoradiograph was obtained through photo-stimulated luminescence, whereby the charge stored by the phosphors is released upon photostimulation and the amount of emitted blue light is directly proportional to the amount of the original radioactive signal. Hence the intensity of the radioactive signal is expressed as photo-stimulated luminescence (PSL) signal. Following autoradiography, the gels were stained as described in Materials and methods (Section 2.3.1) to verify protein levels.

4.3.1 Autokinase assays

4.3.1.1 Time-course assays

BphP1-HK and WT ExsG were incubated in "hot" reaction buffer and the level of phosphorylation was monitored over time. Autoradiography results (Fig. 4.8A and B) confirmed that both proteins possessed autokinase activity. The extent of ^{32}P incorporation increased over time in a linear fashion although in the case of BphP1-HK autophosphorylation rate seemed to decrease after 40 min, possibly due to substrate (protein) limitation. No activity was detected in ExsG control reaction that did not

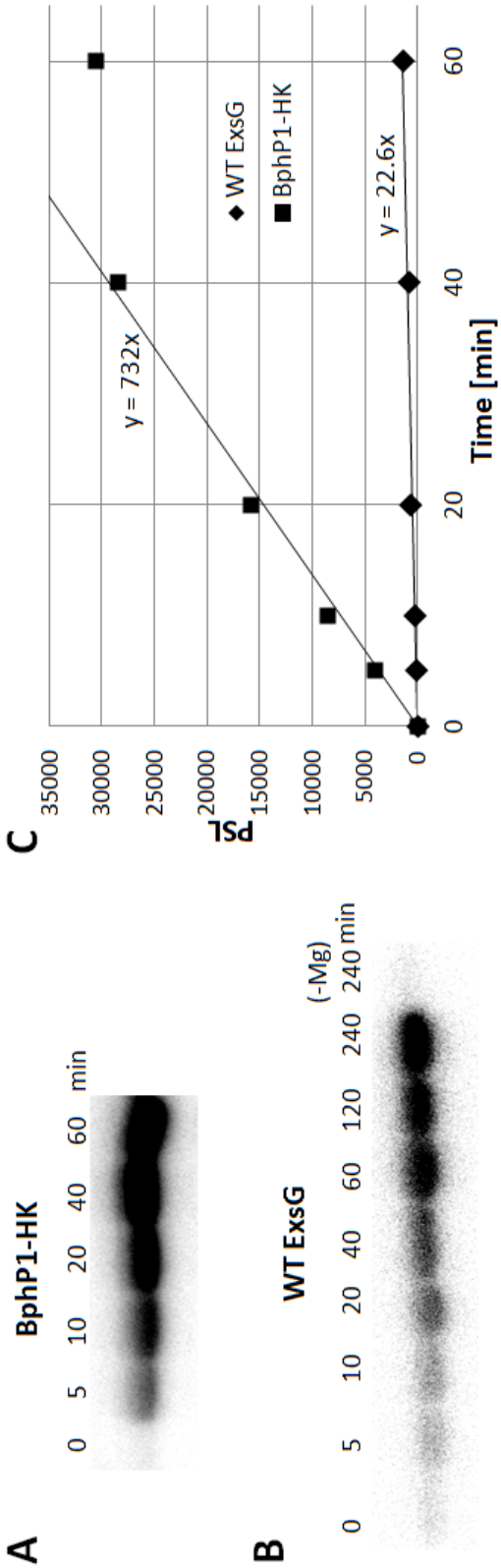


Figure 4.8: Autophosphorylation rates of BphP1-HK and WT ExsG. Autoradiographs showing the results of BphP1-HK (A) and WT ExsG (B) autokinase assays. C - relative autophosphorylation rates; PSL - photostimulated luminescence. Note: BphP1-HK autophosphorylation was monitored over 1 hour only.

contain magnesium chloride (Fig. 4.8B), which confirmed the requirement of divalent cations in the autophosphorylation reaction.

Autokinase assays of WT ExsG and BphP1-HK were performed on different days, using different batches of radioactively labelled ATP and therefore an accurate comparison of autophosphorylation rates for these two proteins could not be made. Nonetheless, the activity of BphP1-HK seemed to be significantly higher than that of ExsG; using BphP1-HK linear data range (up to 40 min, Fig. 4.8C) the relative autophosphorylation rate was estimated to be 732 PSL min^{-1}, or 1464 PSL $min^{-1}pmol^{-1}$, while WT ExsG rate of ^{32}P incorporation was approximately 30-fold lower (45 PSL $min^{-1}pmol^{-1}$).

As mentioned in Section 4.1.2.1, ExsG does not contain His residue of the signature HWE sequence motif. Although it was previously suggested that this residue is required for the activity of the HWE HK core, as its replacement with a lysine residue in *A. tumefaciens* BphP2 resulted in a loss of autophosphorylation [60], it is clear that at least for ExsG the presence of histidine at this location is not necessary for autokinase activity.

4.3.1.2 ATP dependence assay

An attempt was made to determine the dissociation constant of WT ExsG for ATP using isothermal titration calorimetry, but the protein precipitated when stirring was applied; no binding was detected when the stirring option was not used. Therefore an ATP dependence assay was employed to obtain an estimate of the affinity, both for WT ExsG and BphP1-HK. Samples of each protein were incubated for 1 hour in the hot reaction buffer containing different ATP concentrations, ranging from 0.01 to 1 mM. The radioactive signal intensity per protein band was measured and plotted against the ATP concentration (Fig. 4.9). From these preliminary results it was apparent that the relationship between the extent of ^{32}P incorporation by BphP1-HK and ATP concentration was linear at least up to 1 mM ATP, and that the enzyme did not approach saturation even at this high level of ATP. In contrast, in the case of WT ExsG the rates of ^{32}P incorporation appear to follow hyperbolic behaviour with respect to ATP concentration, approaching saturation at 1 mM. However, rates achieved with BphP1 were much higher than that of ExsG. Implications of these kinetic parameters are addressed further in the Conclusions and discussion.

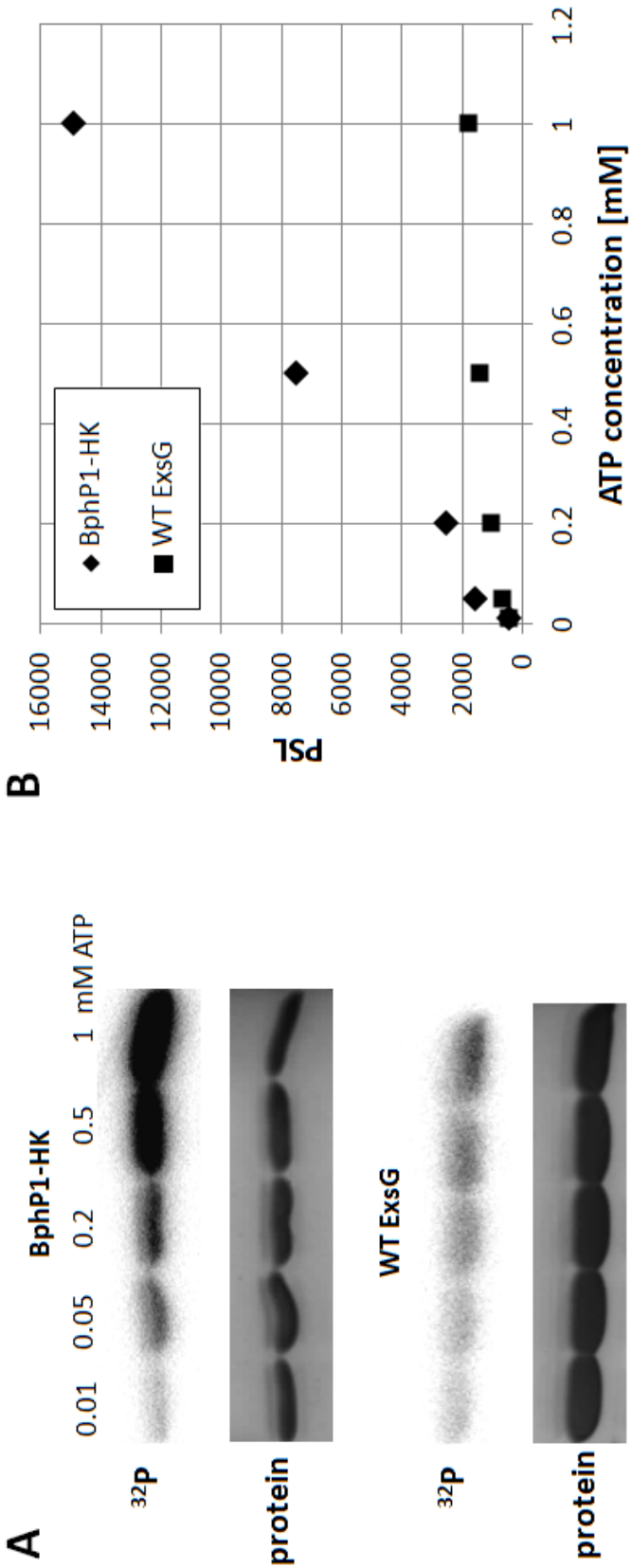


Figure 4.9: Relationship between ATP concentration and BphP1-HK/WT ExsG autophosphorylation extent. A - results of autoradiography. B - plot showing the level of radioactively labelled phosphate incorporation by BphP1-HK and WT ExsG as a function of ATP concentration.

4.3.2 Confirmation of phosphohistidine formation by acid/base stability assay

As discussed in the Introduction (Section 1.1.2.3), phosphohistidine is stable at high pH but acid-labile, in contrast to the properties of O-linked phosphates formed by Ser/Thr/Tyr protein kinases. Therefore acid lability is a characteristic feature that is normally used to confirm the presence of a phosphohistidine. Karniol and Vierstra have previously confirmed the formation of a phosphoramidate bond by HWE kinase cores [60]. Here the acid/base stability test was applied to BphP1-HK and ExsG to confirm the involvement of a histidine residue in the autokinase activity of both proteins.

BphP1-HK and WT ExsG were allowed to autophosphorylate in the "hot" reaction buffer for 2 hours and the level of ^{32}P label associated with the protein was assessed by SDS-PAGE followed by protein transfer onto a membrane and autoradiography. After incubation of membrane strips containing the individual protein bands in either 1 M HCl, 3 M KOH or 50 mM Tris-Cl pH 8.2, the strips were re-assembled and the remaining radioactivity was measured again. The final autoradiograph (Fig. 4.10) demonstrated that the phosphorylation level of each protein was unaffected at high or near-neutral pH, but almost no radioactive signal was detected in the acid-treated protein bands. In conclusion, both BphP1-HK and ExsG form a phosphoramidate intermediate, which is line with the previous studies on canonical and HWE-type kinase cores.

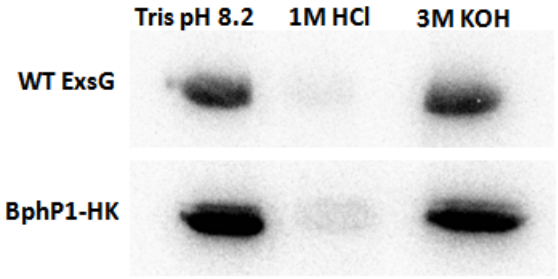


Figure 4.10: Acid/base stability assay of BphP1-HK and WT ExsG phosphoresidues

4.3.3 Temporal stability of the phosphohistidines of BphP1-HK and WT ExsG

To determine if the temporal stability of phosphohistidine intermediates differs between the two HKs, BphP1-HK and WT ExsG were allowed to autophosphorylate in hot reaction buffer for 2 hours, unreacted ATP was removed and the phosphorylated proteins were incubated at room temperature with samples taken at specific time points. Auto-

radiographs (Fig. 4.11) demonstrate that there was no observable decrease in the extent of phosphorylation of either protein over the period of 2 hours. Therefore, the phosphohistidines formed in both proteins are stable over time, at least for 2 hours under the applied conditions. Interestingly, with half-lives longer than 120 min, phosphorylated BphP1-HK and ExsG appear to be more stable than other HKs studied so far - the half-lives of phosphorylated full-length sensor HKs are usually 30 - 90 min [21]. In the case of BphP1-HK phosphohistidine stability may be increased due to the absence of the entire sensory region, while ExsG, as will be discussed later in this chapter, is not a sensor HK, which could explain extended temporal stability of phosphorylation.

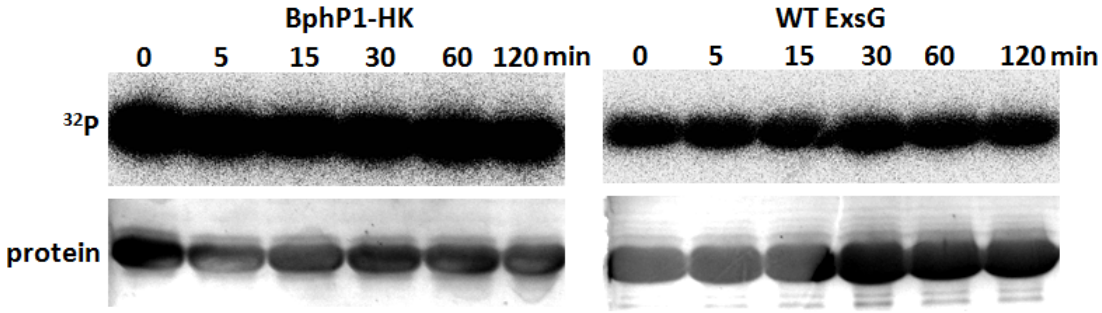


Figure 4.11: Temporal stability of BphP1-HK and ExsG phosphohistidines

4.3.4 Determination of ExsG histidine phosphorylation site

Karniol and Vierstra identified His523 as the phosphorylatable histidine residue within *A. tumefaciens* BphP2 kinase core. Based on the alignment of *At*BphP2 and NT-26 ExsG kinase core amino acid sequences, of which part is shown in Figure 4.12, it seemed likely that ExsG His151 would constitute the site of phosphorylation. However, 41 residues downstream of His151 there was another pair of aligned histidine residues. In order to identify the site of phosphorylation within NT-26 ExsG HWE core, by means of site-directed mutagenesis both histidine residues (His151 and His192) were substituted with asparagine. The purified protein variants, ExsG-H151N and ExsG-H192N, seemed to have the same properties as WT ExsG in terms of solubility as well as oligomeric state - the SEC profiles of the wild-type protein and the two mutants were identical. Interestingly, the H151N variant appeared to have none of the higher molecular weight protein bands which were normally observed above the main ExsG protein band on SDS-PAGE gels (Fig. 4.13); those bands were later identified as acetylated and phosphorylated forms of ExsG (see Chapter 5, Section 5.1). The presence of asparagine within the active site may have in some way affected the efficiency of phosphorylation

and acetylation by *E. coli* enzymes.

Figure 4.12: Part of the sequence alignment of the HWE domains of NT-26 ExsG and *A. tumefaciens* BphP2. The active site histidine of BphP2 (His523) and the putative ExsG phosphorylatable histidine (His151) are indicated by the black arrowhead. The other histidine found at the same respective position in both proteins is indicated by a red arrowhead. ClustalX colouring scheme was applied.

The autokinase activity of the two ExsG histidine mutants was compared to that of the wild-type protein. The three ExsG variants were allowed to autophosphorylate in hot reaction buffer for 1 hour and the extent of phosphorylation was examined using autoradiography. Figure 4.13 shows that substitution of His192 had no effect on autophosphorylation, while mutation of His151 completely abolished autokinase activity, thus confirming that this residue constitutes the phosphorylation site within ExsG kinase core.

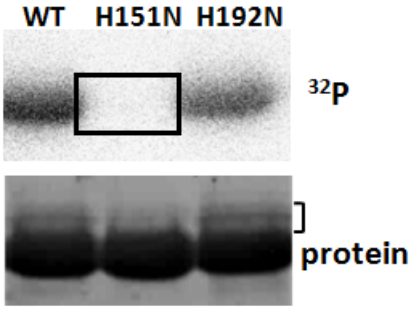


Figure 4.13: Identification of ExsG phosphorylatable histidine residue. Autoradiograph is shown on the top, InstantBlue-stained SDS-PAGE gel is below. The box indicates the expected position of H151N radioactive signal, bracket indicates the area above main ExsG protein band where the additional protein bands are normally observed.

4.4 Determination of the signalling direction of the BphP1-initiated pathway

Transphosphorylation assays were employed to determine all possible phosphotransfer reactions between the proteins belonging to BphP1 gene cluster. These assays involved autophosphorylation of the HK protein (ExsG or BphP1-HK) followed by the addition of the relevant RR protein, such that the RR was present in a molar excess. After

incubation the proteins were separated by SDS-PAGE and the gels were subjected to autoradiography and staining.

4.4.1 Verification of ExsG receiver domain activity

In order to establish if the N-terminal domain of ExsG is a functional receiver domain, transphosphorylation assays as well as phosphoprotein affinity gel electrophoresis were employed.

4.4.1.1 Test of BphP1-ExsG phosphotransfer

Given the presence of the N-terminal receiver domain, ExsG seemed likely to act as a RR. Commonly, transphosphorylation assays are carried out in the presence of ATP to support phosphotransfer. However, if ATP was present in the reaction mixture containing BphP1-HK and ExsG, both proteins would be engaging in autophosphorylation. Therefore, chromatography spin columns were used to remove the unreacted ATP from BphP1-HK autokinase reaction prior to addition of WT ExsG. Phosphorylated BphP1-HK was then incubated with WT ExsG at a molar ratio 1:4 and samples were taken after 5, 30 and 60 minutes of the reaction. Phosphotransfer was evident from the gradual transfer of radioactive signal from BphP1-HK onto ExsG (Fig. 4.14). The appearance of the radioactive signal within ExsG protein band was unlikely to represent ExsG autokinase activity as it was accompanied by the loss of radioactivity from BphP1-HK band. Hence it was concluded that BphP1 phosphorylates the receiver domain of ExsG.

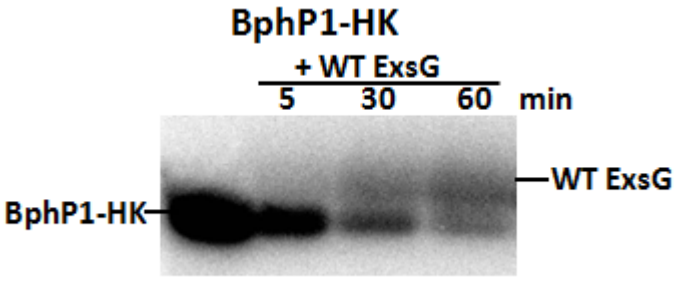


Figure 4.14: WT ExsG phosphorylation by BphP1-HK. The first lane contains phosphorylated BphP1-HK, the other lanes contain both proteins. WT-ExsG - 37.6 kDa, BphP1-HK - 27 kDa.

4.4.1.2 Identification of the phosphorylation site within ExsG receiver domain

Phosphorylatable aspartate is normally located directly after the third β-strand and, as mentioned in Sections 4.1.2.2 and 4.1.3, Asp62 was the most likely site of phosphorylation within ExsG receiver domain. However, there were several other aspartate residues positioned in the vicinity of Asp62, at amino acid positions 55, 57, 67 and 69. To identify the aspartate which receives the phosphoryl group from BphP1-HK, the residue most likely to constitute the site of phosphorylation (Asp62) and two aspartates closest to Asp62 (Asp57 and Asp67) were substituted with asparagines. The positions of the mutated Asp residues in the context of the threaded structure of ExsG receiver domain are shown in Fig. 4.15A. The resulting protein variants did not differ from WT ExsG in terms of solubility, yield or SEC profile, and they retained autokinase activity as shown in Figure 4.15B.

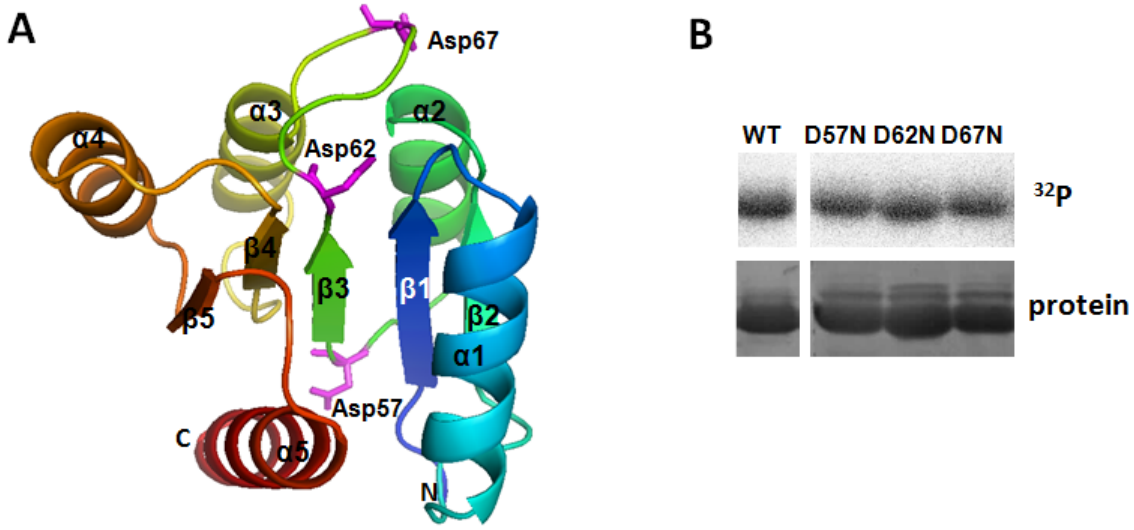


Figure 4.15: Location of the mutated aspartate residues within ExsG receiver domain and autokinase activity of the protein variants. A - the structure of the receiver domain of ExsG threaded onto CheY (PDB ID 3CHY); the positions of the three mutated aspartate residues are indicated (magenta). B - autokinase activity of the three ExsG site-specific mutants with respect to wild-type protein.

The ability of these aspartate mutants to receive phosphoryl group from BphP1-HK was assayed using transphosphorylation assays which also included WT ExsG and the two histidine mutants described in Section 4.3.4. As all proteins (except for H151N variant) possessed autokinase activity, unreacted ATP was removed from BphP1-HK autophosphorylation reaction before setting up transphosphorylation reactions. After 1 hour incubation of phosphorylated BphP1-HK with each separate ExsG variant (molar

ratio 1:4 in each reaction) it was evident that ExsG-D62N could not be phosphorylated on the receiver domain, while all other ExsG protein products were capable of receiving the phosphoryl group (Fig. 4.16A). Therefore, as predicted, Asp62 constitutes the phosphorylation site within ExsG receiver domain.

Further confirmation of Asp62 involvement was provided by using small-molecule phosphate donor synthesised in-house, phosphoramidate (PA), to phosphorylate WT ExsG and ExsG-D62N. To separate the phosphorylated forms, the proteins were subjected to phosphoprotein affinity gel electrophoresis which utilises Phos-tag^{TM}/Mn^{2+} moiety to bind phosphoryl groups and thus retard their migration rate through the gel. After 2 hours of incubation with PA majority of WT ExsG molecules were present in the upshifted band, corresponding to phosphorylated protein, while ExsG-D62N molecules were retained in the lower protein band (Fig. 4.16B). In conclusion, both BphP1 and PA specifically phosphorylate ExsG on Asp62.

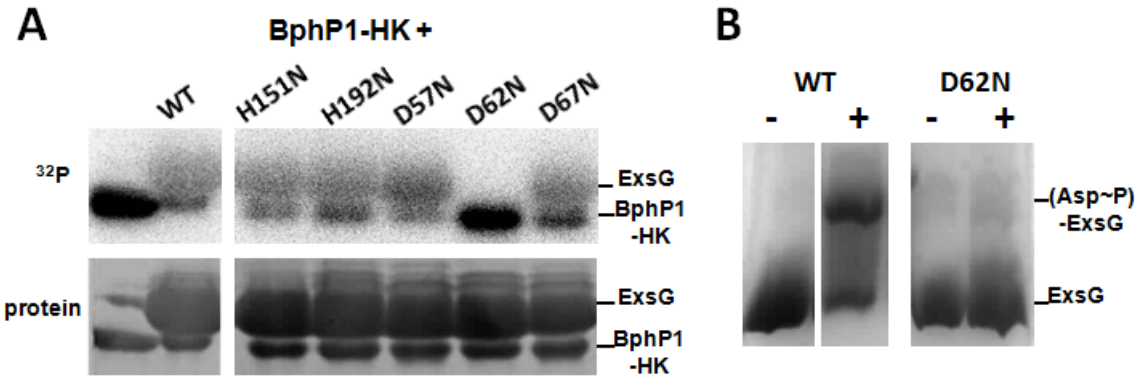


Figure 4.16: Identification of the phosphorylatable aspartate within ExsG receiver domain. A - results of transphosphorylation assays; the first lane contains phospho-BphP1-HK, the other lanes contain BphP1-HK and the specified ExsG variant. B - Coomassie-stained Phos-tag^{TM}/Mn^{2+} SDS-PAGE gel (7.5% acrylamide) showing the extent of phosphorylation of WT ExsG and D62N after 2 hours incubation with PA.

Even though two ExsG aspartate mutants were phosphorylated by BphP1-HK, the phosphotransfer efficiency of ExsG-D67N variant appeared to be reduced. Radioactive signal associated with BphP1-HK did not seem to decrease as dramatically as in the other transphosphorylation reactions. The experiment was repeated using all ExsG variants, except for H151N, and the resulting autoradiograph confirmed that phosphotransfer onto D67N variant was impaired (Fig. 4.17). One explanation for the observed phenomenon is that Asp67, which is located within the same loop region as the phosphorylatable residue, is involved in the phosphotransfer reaction. However, it is apparent from Figure 4.17 that the total amount of radioactivity retained by the proteins in this transphosphorylation reaction is significantly higher than in the other

reactions. A possible scenario that would fit this observation is that Asp67 participates in autodephosphorylation reaction and accelerates the rate of phosphoaspartate hydrolysis. Without continuous autodephosphorylation, and if the equilibrium of the phosphotransfer reaction favours phosphorylated BphP1-HK, the amounts of resulting phosphoaspartate would be reduced.

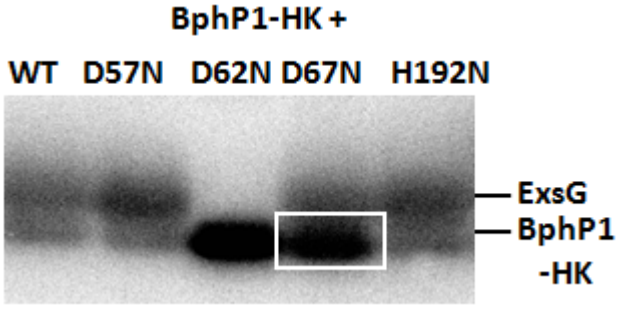


Figure 4.17: Effect of ExsG Asp67 substitution on BphP1-ExsG phosphotransfer efficiency. The box indicates the position of the radioactive signal associated with BphP1-HK in trans-phosphorylation reaction with ExsG-D67N.

4.4.1.3 Test of ExsG receiver domain phosphorylation by ExsG kinase core

Given that ExsG receiver domain was successfully phosphorylated by the kinase core of BphP1, it seemed plausible that the kinase core of ExsG may also phosphorylate its own receiver domain (see Fig 4.18 for a schematic diagram). To verify this possibility, ExsG-REC was added in approximately 4-fold molar excess to WT ExsG, which was previously allowed to autophosphorylate for 1 hour, and the proteins were incubated together for another hour. Figure 4.18 shows that no phosphotransfer onto ExsG-REC occurred, which rules out the possibility of phosphotransfer between the kinase core located on one ExsG molecule and the receiver domain of another ExsG molecule.

4.4.2 Determination of cognate pairs of histidine kinases and response regulators by phosphotransfer profiling

The technique termed phosphotransfer profiling allows to discriminate specific phosphotransfer from cross-phosphorylation even with only two time points, for instance 1 and 60 minutes. It has been shown by Skerker and co-workers [89] that HKs can cross-phosphorylate non-cognate RRs but with very low efficiency and over prolonged incubation, while cognate RRs would be phosphorylated rapidly, often within a minute.

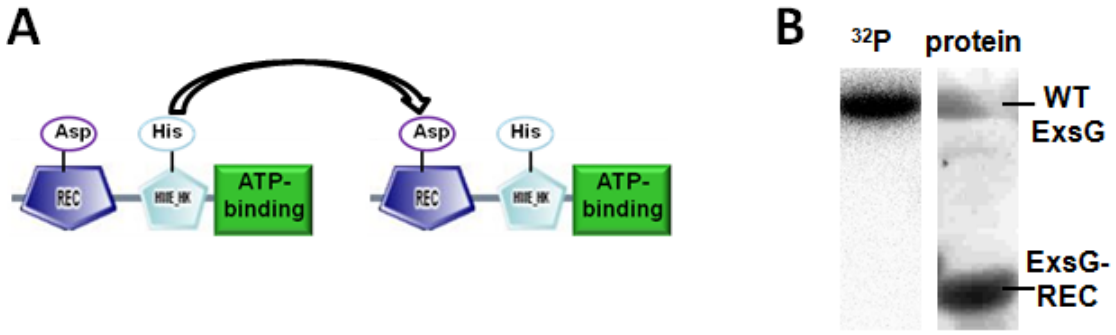


Figure 4.18: Test of intermolecular phosphotransfer between ExsG kinase core and receiver domain. A - schematic diagram of intermolecular phosphotransfer. B - autoradiograph and the corresponding SDS-PAGE gel.

Phosphotransfer can be visualised by monitoring the increase of radioactive signal over the RR protein band, and the decrease in the extent of HK phosphorylation.

4.4.2.1 Phosphotransfer profiling with BphP1-HK

The phosphotransfer from BphP1 to AgR was previously reported in *A. tumefaciens* [71] and the ability of BphP1-HK to phosphorylate WT ExsG in NT-26 was demonstrated earlier in this section. To confirm the occurrence of BphP1-AgR phosphotransfer in NT-26, and to verify if both AgR and ExsG constitute cognate RRs of BphP1, phosphotransfer profiling with two time points (5 and 60 minutes) was employed. The fourth protein belonging to BphP1 gene cluster, ExsF, was also tested to establish if it too can be phosphorylated by BphP1. ExsG-H151N variant was used instead of WT ExsG to eliminate the signal arising from ExsG autokinase activity. As demonstrated in Section 4.4.1.3, stand-alone ExsG receiver domain could not be phosphorylated by ExsG kinase core, and therefore ExsG-REC construct was included herein to test if the receiver domain of ExsG retains activity as a stand-alone domain protein.

The results demonstrate that BphP1-HK was capable of phosphorylating each of the four RR proteins tested, but there was a clear difference in the rate of phosphotransfer onto ExsF and all other RRs (Fig. 4.19). AgR, ExsG-H151N and ExsG-REC were efficiently phosphorylated within the first 5 minutes of the reaction, while phosphotransfer onto ExsF was significantly slower and became apparent only after prolonged incubation, suggesting that this protein is not a cognate RR of BphP1 and would not be phosphorylated by BphP1 *in vivo*. As discussed in the Introduction (Section 1.3.5), the occurrence of such cross-talk to non-cognate RRs *in vitro* is most likely due to a high level of sequence and structure similarity between receiver domains.

The level of phosphorylation achieved by ExsG-REC was comparable to that of

ExsG-H151N. The observed phosphotransfer onto ExsG-REC implies that ExsG receiver domain retained activity in the absence of the kinase core.

Notably, phosphorylation of both ExsG constructs by BphP1-HK seemed to occur at a similar rate to AgR phosphorylation, which demonstrated that both ExsG and AgR are the cognate RRs of BphP1. Therefore, the signalling pathway downstream of BphP1 is intrinsically branched.

4.4.2.2 Phosphotransfer profiling with ExsG

ExsG was shown to be an active HK, and it was therefore likely that AgR or ExsF, or both these proteins, could be phosphorylated by ExsG. As its autokinase activity seemed to be lower than that of BphP1-HK (see Section 4.3.1), the wild-type protein was allowed to autophosphorylate for a longer period of time (1 hour as opposed to 30 min for BphP1-HK autophosphorylation). AgR and ExsF were present at an approximately 3-fold excess, unlike in BphP1-HK phosphotransfer profiling where RRs were in over 6-fold molar excess.

Figure 4.19 shows that no phosphotransfer occurred between ExsG and AgR, even after 1 hour, while ExsF was efficiently phosphorylated, even though the level of ExsF phosphorylation achieved at the two time points was significantly different. Prolonged incubation resulted in a dramatic increase in the radioactive signal, which would suggest that the phosphotransfer efficiency is lower than that observed in BphP1-ExsG and BphP1-AgR transphosphorylation reactions. However, after 5 minutes of incubation of ExsG with ExsF there was almost no radioactivity confined to ExsG protein band; decrease in HK phosphorylation level was implied by Skerker and co-workers as a sign of specific phosphorylation [89]. Furthermore, considering the very low level of phosphotransfer between BphP1-HK and ExsF, it was evident that ExsF constitutes the cognate RR of ExsG.

4.4.3 Determination of the phosphotransfer preference of BphP1

The fact that BphP1 efficiently phosphorylates both ExsG and AgR in separate transphosphorylation reactions raised the question concerning phosphorylation preference *in vivo*. To determine if BphP1 exhibits a phosphotransfer preference for either of the two proteins, a preliminary preferential assay was set up, whereby phosphorylated BphP1-HK was incubated simultaneously with ExsG-H151N and AgR. The two RRs were present at approximately 5-fold molar excess over BphP1-HK. As shown in Figure 4.20, the increase in AgR phosphorylation level was significantly slower than in the

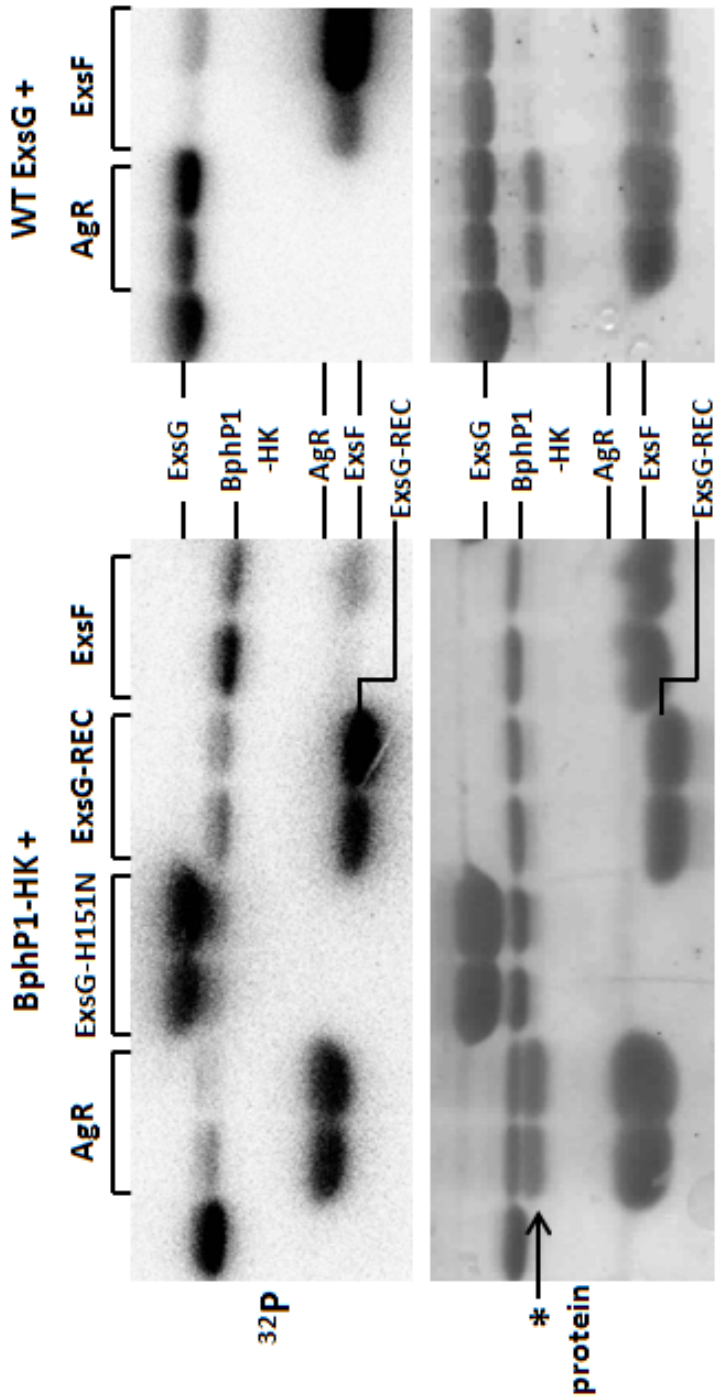


Figure 4.19: Determination of all phosphorylation substrates of BphP1 and ExsG using phosphotransfer profiling. First lane on each autoradiograph/gel represents the respective autokinase reaction. Each of the two lanes represents the respective transphosphorylation reaction at 5 and 60 min. Asterisk denotes an AgR-associated contaminant.

case of ExsG, and the latter protein appeared to be efficiently phosphorylated within the first few minutes of the reaction. This observed phosphorylation preference was reproducible, and it was concluded that the preferred substrate of BphP1 was ExsG. Therefore, BphP1 affinity for ExsG may be higher than the affinity for AgR.

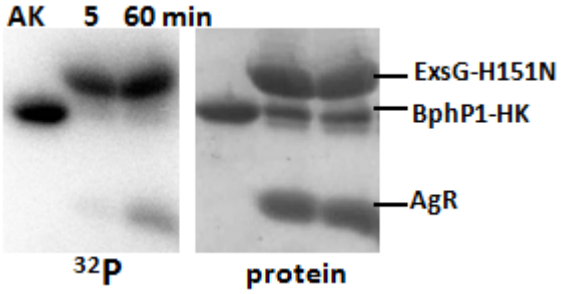


Figure 4.20: Phosphotransfer preference assay of BphP1-HK. AK - BphP1-HK autokinase reaction; 5 and 60 min refer to the incubation time of all three proteins.

4.5 Attempts to determine the response mediated by AgR and ExsF

The direction of the BphP1-initiated signalling pathway seems to be linked to the arrangement of genes forming the BphP1 cluster (summary in Fig. 4.21), and it was demonstrated that ExsG phosphorylates ExsF, despite the fact that ExsF is encoded in the opposite direction. ExsG was later shown to act as a RR that phosphorylates ExsF when BphP1 autophosphorylation is switched off (see Chapter 6). Therefore the two stand-alone receiver domain proteins, AgR and ExsF, most probably constitute the final phosphoreceptors in this signalling system, and it is likely that they would not coexist in the phosphorylated form. Such single-domain RRs are usually involved in chemotaxis or phosphorelay [54]; several approaches were undertaken to determine if these two proteins could be involved in either of the two processes.

4.5.1 Test for phosphotransfer onto histidine phosphotransfer domain

There was only one gene encoding a histidine phosphotransfer (Hpt) domain protein (13 kDa) identified in NT-26 genome. It was cloned and initially transformed into *E. coli* BL21 (DE3) expressing strain. The pI predicted for this protein using the ProtParam

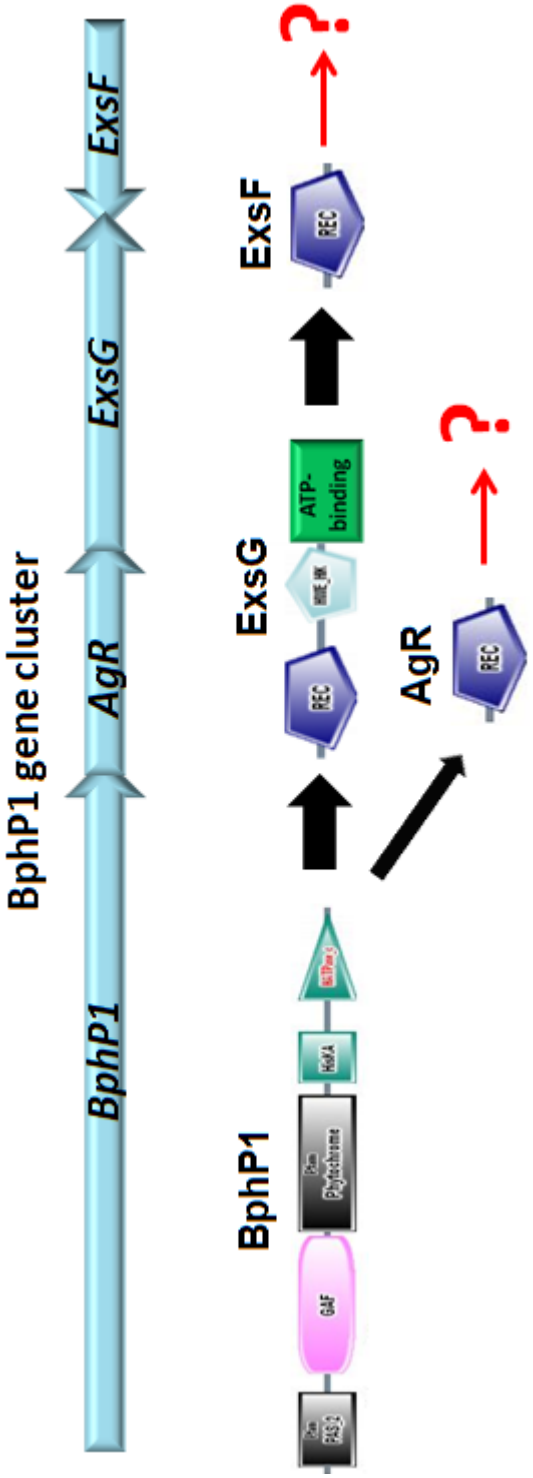


Figure 4.21: Summary of the signalling direction downstream of BphP1. The thickness of black arrows correlates with phosphotransfer preference. Note: the phosphorylation-dependent regulation of ExsG HK core activity by the receiver domain is not indicated; it is described in Chapter 6.

online tool[1] [148] was 8.1, which is unusually high for a cytoplasmic protein; with the N-terminal purification tag the predicted pI was 7.2. Given that the standard buffers used in protein purification herein were of pH 8.2, attempts were made to purify this protein in 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6 buffer containing 0.5 or 1 M NaCl, 1 mM β-ME and 15% glycerol. However, the yield of purified soluble protein was very low (<0.1 mg per gram of cells) and almost all of the protein was retained in the insoluble fraction.

The vector containing Hpt gene was therefore introduced into two different types of *E. coli* ArcticExpress™ cells, which are designed for expression of proteins at lower temperatures. The expression was tested both with and without the addition of gentamycin to the starter culture, and the protein was expressed for 24 hours at 16°C. The protein purified from an ArcticExpress™ cell pellet was still mainly found in the insoluble fraction but the yield from NiNTA purification was higher than before (0.1 - 0.2 mg per gram of cells, Fig. 4.22A). As the pI of BphP1-HK was 6.2 and the protein precipitated out of solution at pH below 7, Hpt protein was buffer-exchanged into 50 mM Tris-Cl pH 7, 0.5 M NaCl, 1 mM β-ME, 15% glycerol buffer following tag removal with TEV protease. However, Hpt without the purification tag could not be concentrated to more than ∼0.3 mg/ml, and it was associated with numerous contaminants (Fig. 4.22A, right panel).

Another approach was therefore applied, whereby the protein was purified from ArcticExpress™ cell pellet in 50 mM Tris-Cl pH 8.5, 0.5 M NaCl, 1 mM β-ME, 15% glycerol buffer and the tag was retained, so that the pI of the protein should remain at 7.2. The yield from NiNTA purification was not significantly better than at pH 6 (Fig. 4.22B), but the purity was much higher, possibly due to the fact that at alkaline pH the polyhistidine tag binds to nickel ions with a higher affinity. The maximum concentration of the purified Hpt protein with the retained tag was ∼0.4 mg/ml. BphP1-HK/AgR and WT ExsG/AgR transphosphorylation reactions were incubated for 1 hour, then Hpt was added to half of each reaction volume; the remaining part of each reaction was stopped by adding the loading dye. After 1 hour NiNTA beads were added to the Hpt-containing reactions for about 10 minutes to pull out the tagged Hpt protein. Hpt was visible on the SDS-PAGE gel in the beads fraction, but it did not seem to have any associated radioactivity in either of the two transphosphorylation reactions (Fig. 4.22C). These results suggested that AgR and ExsF are unlikely to transfer the phosphoryl group onto this Hpt domain protein, but this could also be due to the inherent insolubility of Hpt. Therefore further experiments would be necessary to

[1] http://web.expasy.org/protparam/

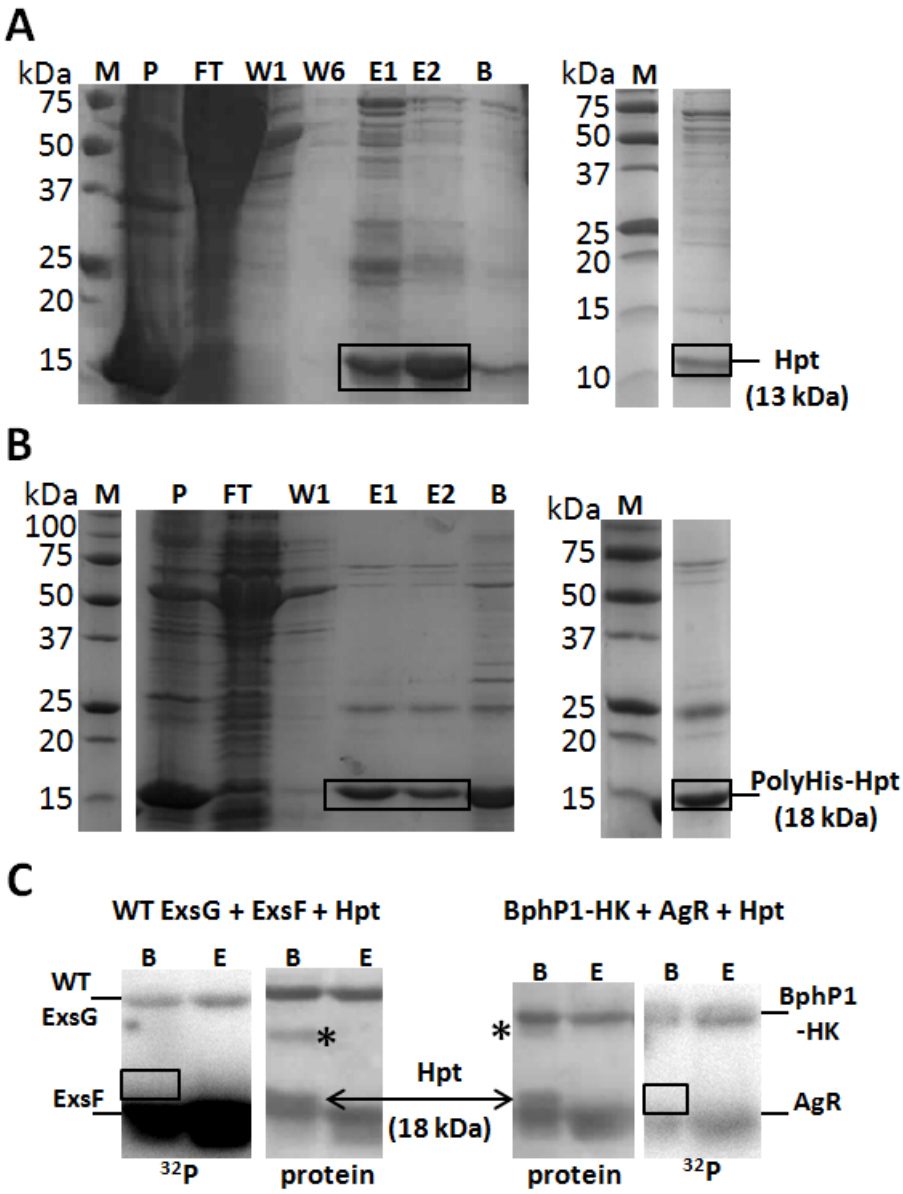


Figure 4.22: Purification of Hpt domain protein and phosphotransfer test. SDS-PAGE analysis of NiNTA purification at pH 6 (A) and 8.5 (B). The concentrated protein after tag removal (A) and with the tag retained (B) is shown on the right. M - marker, P - pellet (insoluble fraction), FT -flowthrough, E - elution, W- wash, B - beads fraction. C - phosphotransfer test on the tagged form of Hpt at pH 8.5. Following 2 hours incubation of Hpt with each transphosphorylation reaction NiNTA beads were added for 10 min and the reactions were centrifuged briefly. B - beads fraction, E - supernatant. Position of Hpt protein band is indicated by the double arrow. The boxes indicate the expected position of Hpt-associated radioactive signal. Asterisk denotes a contaminant associated with Hpt.

exclude the possibility that signalling downstream of BphP1 involves a phosphorelay.

4.5.2 Determination of the likely timescale of the response

The stability of ExsG, ExsF and AgR phosphoaspartates was examined in the presence, as well as in the absence, of their cognate HKs, in an attempt to establish whether the output triggered by BphP1-initiated signalling cascade is likely to involve a rapid response, as observed in chemotaxis, or a gradual one, such as a change in gene expression.

4.5.2.1 Determination of AgR, ExsF and ExsG phosphoaspartate stability

Receiver domains frequently catalyse their own dephosphorylation [43, 3, 4], and the rate of this process depends on the requirements of the particular TCS. In the case of chemotaxis, autodephosphorylation of CheY occurs rapidly, with phosphoaspartate half-life usually under 1 minute [149, 150, 44]. On the other hand, TCSs which do not need to respond instantaneously to environmental changes, for instance those involved in regulating gene expression, may involve phosphoaspartates which remain stable over hours or even days [4, 43].

To determine if the TCS studied here is likely to evoke a rapid type of response, as in the case of chemotaxis, AgR, ExsF and wild-type ExsG, as well as ExsG receiver domain purified separately (ExsG-REC), were phosphorylated using a small-molecule phosphate donor and, after removing unreacted compound, they were incubated in the presence of magnesium chloride at room temperature. PA was used as the phosphate donor for all proteins, but ExsF was also assayed separately after phosphorylation with acetyl phosphate. Samples taken at several time points were analysed by phosphoprotein affinity gel electrophoresis (Fig. 4.23A). The relative amount of phosphorylated (upshifted band) and unphosphorylated protein (bottom band) at each time point was estimated based on the staining intensity and normalised. The starting amount of phosphorylated protein (at 1 minute) was taken as 100% phosphorylation level.

As evident from the plot of phosphorylation level versus time (Fig. 4.23B), all phosphoaspartate residues exhibited a half-life of over 90 min, which suggested that they do not undergo rapid autodephosphorylation. The phosphoaspartate of WT ExsG seemed to be even more stable over time, with only 10% decrease in the phosphorylation extent after 2 hours. Hence it seemed unlikely that this TCS would evoke a response that requires tight regulation of RR phosphorylation level.

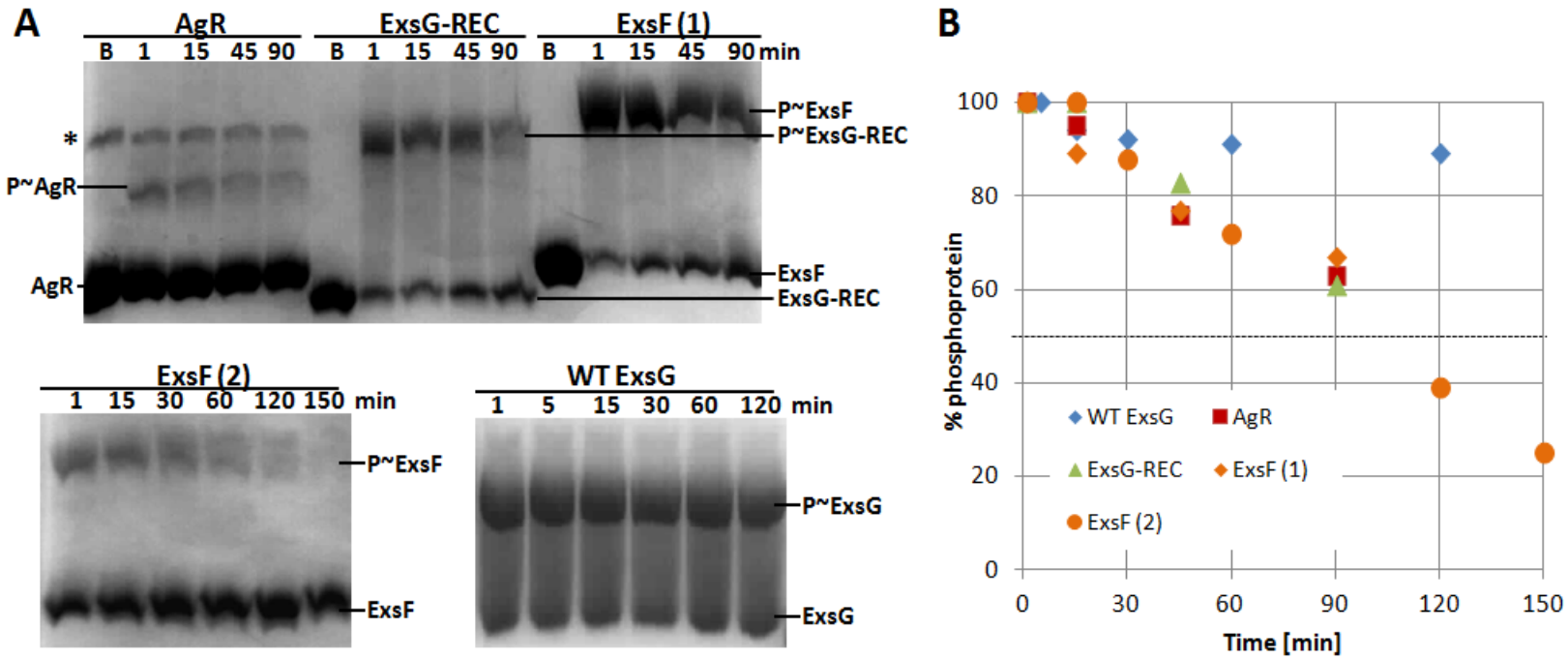


Figure 4.23: Determination of AgR, ExsF and ExsG-REC phosphoaspartate stability. A - analyses of phosphoprotein affinity gel electrophoresis of phosphorylated RR proteins. ExsF was phosphorylated either with PA (1) or with acetyl phosphate (2). Asterisk denotes a contaminant associated with AgR. B - plot of the change in proportion of phosphorylated RRs against time. The dashed line indicates 50% phosphorylation level.

4.5.2.2 Phosphatase activity of ExsG and BphP1

The phosphoaspartates of AgR, ExsF and ExsG were shown in the previous section to be relatively stable over time. However, the phosphorylation levels of RRs can also be negatively regulated by their cognate HKs. As discussed in the Introduction (Section 1.1.2.4), phosphatase activity of HKs is a common feature of many TCSs, and ADP has been shown to stimulate that activity in at least some HKs. An attempt was made in this project to establish if the two HK proteins within BphP1 signalling system - BphP1 and ExsG - are capable of negatively regulating the phosphorylation levels of their cognate RRs. Rapid dephosphorylation mediated by the HKs would suggest that this TCS requires an instantaneous stimulus-dependent adjustment of RR phosphorylation level.

According to Willett and Kirby [147], a conserved asparagine or threonine residue located 4 residues downstream of the phosphorylatable histidine is necessary for phosphatase activity. Both NT-26 and *A. tumefaciens* BphP1 contain an alanine residue at this position (see Fig. 4.1), suggesting that BphP1 is unlikely to exhibit phosphatase activity. In contrast, both ExsG sequences featured an asparagine (Asn155 in NT-26 ExsG, see Fig. 4.2), but, given that the overall sequence divergence of HWE kinase cores, the presence of that residue could be unrelated to phosphatase activity.

ExsG phosphatase assay involved the cognate RR, ExsF, while WT ExsG was used to assay the phosphatase activity of BphP1-HK; the possible negative regulation of AgR phosphorylation by BphP1 has not been addressed here. BphP1-HK was allowed to autophosphorylate, purified from unreacted nucleotide and incubated with WT ExsG for 1 hour to allow phosphotransfer. ExsG and ExsF were incubated together for 1 hour in the hot buffer, after which unreacted ATP was removed. ADP was added to half of each reaction volume and the reactions (with and without ADP) were incubated for 2 hours, with samples taken at specific time points. Following electrophoresis and autoradiography, the intensity of the radioactive signal associated with each RR (WT ExsG or ExsF) was quantified and the extent of phosphorylation was normalised to the starting (100%) phosphorylation level. The half-life was estimated as the time at which ∼50% of the protein remains phosphorylated based on the plot of phosphorylation decay versus time. Autoradiographs and the plots of phosphorylation level versus time (Fig. 4.24) showed that the rate of hydrolysis of ExsF phosphoaspartate (half-life ∼30 min) was significantly higher than that of ExsG phosphoaspartate (half-life ∼120 min).

As residual ATP from preceding transphosphorylation reactions could affect phosphatase reactions, another approach was also undertaken to determine if unphosphorylated BphP1-HK can promote the hydrolysis of ExsG phosphoaspartate. ExsG

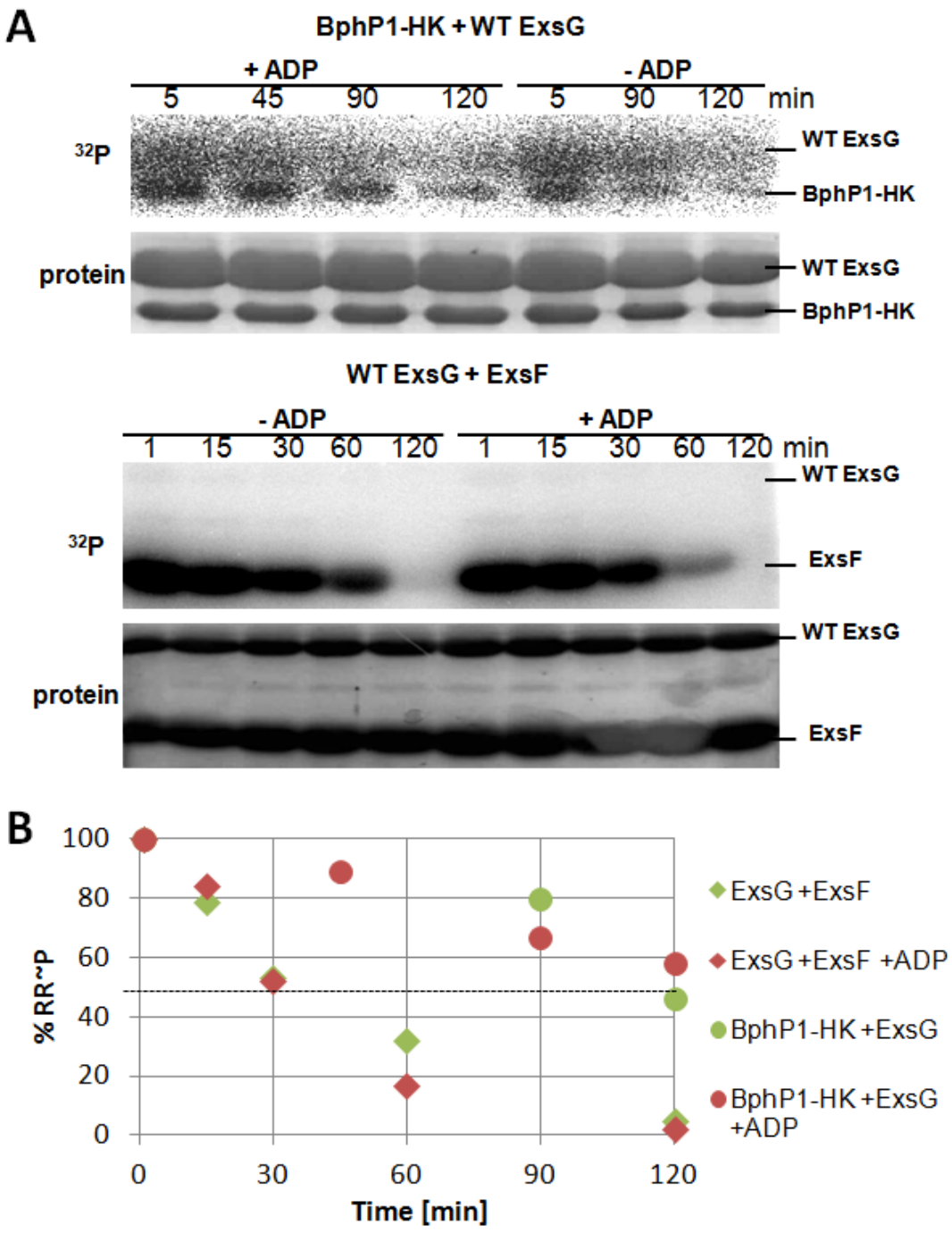


Figure 4.24: Radioactive phosphatase assays of BphP1-HK and WT ExsG. A - autoradiographs and corresponding SDS-PAGE gels of the respective reactions. B - plot showing the decay of phosphorylated RRs with time. Dashed line indicates 50% RR phosphorylation level.

was phosphorylated by PA over 2 hours and after unreacted compound was removed, BphP1-HK and ADP were added. Samples taken at specific time points were analysed by phosphoprotein affinity gel electrophoresis (Fig. 4.25). In agreement with the results from the radioactive assay described above, it appeared that the level of ExsG phosphorylation, represented by the upshifted band, remained unchanged over the course of 30 minutes. This type of assay could not be applied to ExsG/ExsF reaction as phosphorylated form of ExsF partly overlapped with ExsG protein band.

When compared to the rate of RR autodephosphorylation (see Section 4.5.2.1), the rate of phosphate hydrolysis occurred faster in the presence of the cognate kinase. The shortest phosphoaspartate half-life was determined to be 30 min, in the case of ExsG/ExsF phosphatase assay, and the rate of ExsF phosphoaspartate decay was increased 3-fold by the presence of the cognate HK. The half-life of ExsG phosphoaspartate was not determined as it exceeded 2 hours, but after the same incubation time in the presence of BphP1-HK the phosphorylation level of ExsG decreased to ∼50%. Such differences are, however, insignificant in comparison with the usually dramatic effect of the addition of unphosphorylated HK to its cognate phosphorylated RR. For instance, the phosphoaspartates of two RRs belonging to a branched TCS in *Synechococcus elongatus* were stable over hours, but addition of the cognate HK and ADP reduced their half-lives to less than 5 minutes [93]. Similarly, *Myxococcus xanthus* CrdS diminished the phosphorylation level of its cognate RR to 2% within just 5 minutes [147]. Furthermore, the presence of ADP did not seem to stimulate the phosphatase activity of either HK tested here. Therefore neither BphP1 nor ExsG seemed to exhibit significant phosphatase activity towards the cognate RRs tested herein, at least under the applied conditions.

Taking into consideration the temporal stability of all phosphoaspartate residues tested herein, both in the presence and in the absence of their cognate HKs, BphP1 signalling cascade seemed likely to affect gene expression rather than evoke a rapid type of response, such as reversing the direction of flagellar motor rotation. Similar dynamics characterise two cyanobacterial phytochrome systems in *Calothrix* PCC7601, where neither of the two phytochromes exhibited phosphatase activity towards the cognate single-domain RR, and the deduced autodephosphorylation rate of the RRs was low [151].

Lack of BphP1 phosphatase activity would be in line with the absence of the highly conserved Asn/Thr residue in the vicinity of active site histidine, while the respective asparagine within the HWE core of ExsG (Asn155) may play a different role. However, in the case of BphP1-HK the absence of the N-terminal region could potentially result

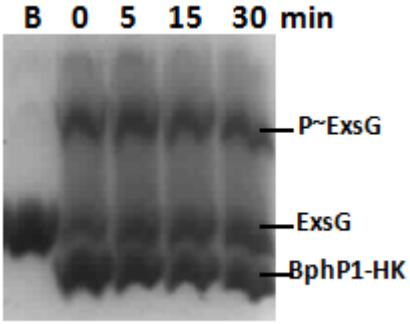


Figure 4.25: Non-radioactive assay of BphP1-HK phosphatase activity towards ExsG. Time was measured since the addition of BphP1-HK. B - ExsG before incubation with PA.

in reduced phosphatase activity, as observed in the case of NblS lacking the preceding HAMP domain [93]. The putative involvement of the sensory region in regulating phosphatase activity has been postulated based on the crystal structure of ThkA (see Introduction, Section 1.1.2.1 and [37]). Therefore full-length BphP1 would need to be assayed to verify if it exhibits a significant level of phosphatase activity.

Notably, the preliminary results presented above do not exclude the possibility that the phosphorylation levels of ExsF and ExsG are regulated by accessory proteins encoded elsewhere in the genome. Furthermore, the experimental setup applied here may be flawed as residual ATP could still be present in the reaction; even though ExsG dephosphorylation rates appeared to be similar in both radioactive and non-radioactive assays, the experiment would ideally be conducted using a radioactively labelled phosphate donor to phosphorylate the RRs prior to the addition of unphosphorylated HK.

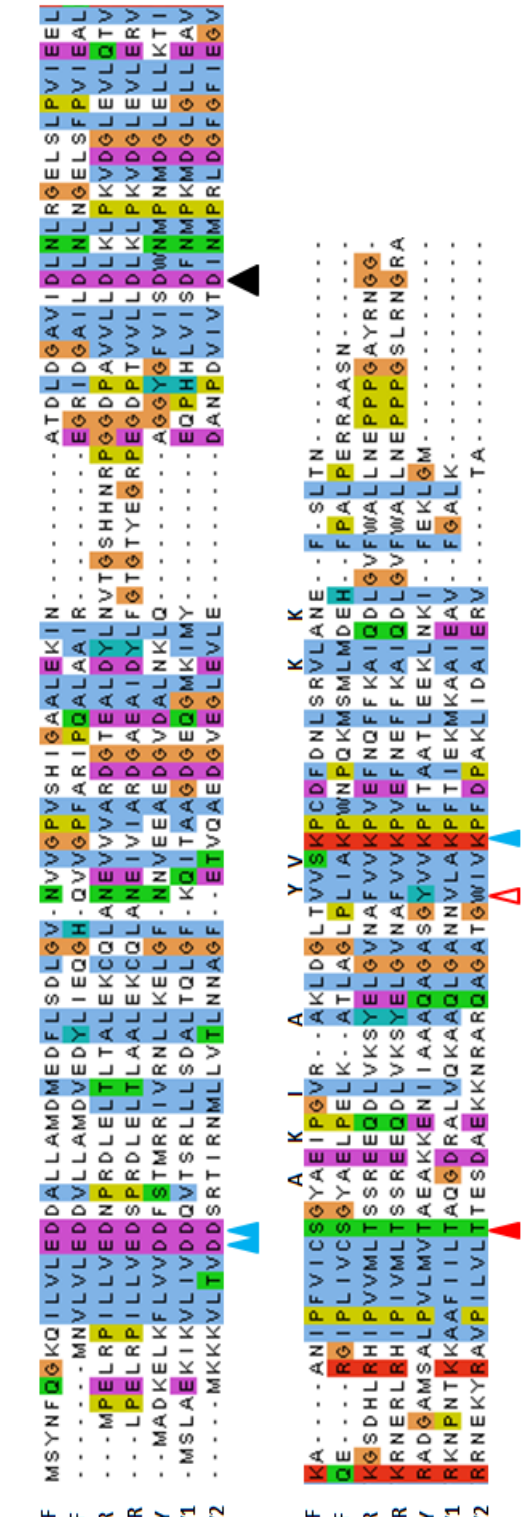
4.5.3 Comparison of AgR and ExsF to CheY

CheY has been extensively characterised and three-dimensional structures have been determined of its inactive [152, 72] and activated conformation [153], as well as of the complex of activated CheY and the N-terminal portion of the flagellar motor switch protein, FliM [154]. The activated form refers to CheY bound to berylloﬂuoride, a phosphate mimic. Based on these structures the transition from inactive to active state as well as the mode of FliM peptide binding have been elucidated. Phosphorylation seems to induce relatively modest structural perturbations compared to the inactive form of the protein [153]. Thr87 forms a hydrogen bond with one of the fluorines which displaces strand β4, and this in turn reorients the side chain of Tyr106. Burial of Tyr106 side chain affects the surface of the α4/β5/α5 region and stabilises helix α4, while the hydrogen bond formed between tyrosine hydroxyl group and Glu89 stabilises

the β4/α4 loop. It appears that different receiver domains display varying magnitudes of phosphorylation-induced perturbations, but the repositioning of the conserved threonine and tyrosine residues ("Y-T coupling") has been reported to be the key phenomenon accompanying activation [4].

The sequences of AgR and ExsF were compared to both *E. coli* and NT-26 CheY proteins. As discussed in Chapter 3, NT-26 CheY1 and CheY2 both contain the essential active site residues and are thus likely to be functional, but their sequences are only 35% identical. As the functions of the two CheY proteins could not be determined, both sequences, together with *E. coli* CheY, were used in the construction of the multiple sequence alignment with AgR and ExsF sequences from NT-26 and *A. tumefaciens* (Fig. 4.26). The residues corresponding to *E. coli* CheY residues shown to interact with FliM peptide [154] have mostly different identities in the other sequences, including both NT-26 CheY proteins. Furthermore, NT-26 CheY proteins, AgR and ExsF also exhibit substantial variability at the respective positions, suggesting that AgR and ExsF may not act as CheY. It seemed likely that the *E. coli* CheY Tyr106, involved in Y-T coupling, may be substituted with the other two aromatic residues, phenylalanine or tryptophan, as both NT-26 CheY2 and AgR do not contain a tyrosine at that position. Interestingly, however, both ExsF and CheY1 contain a small hydrophobic amino acid (Val or Leu) instead of the aromatic residue.

Given that NT-26 CheY proteins were significantly different from *E. coli* CheY at the positions implied in FliM binding, the N-terminal portion of NT-26 and *A. tumefaciens* FliM proteins were compared against *E. coli* FliM (UniProt entry P06974) (Fig. 4.27). As reported by Lee and co-workers, there are several highly conserved residues within the first 16 residues of FliM which interact with CheY, and these residues appeared invariant across three different bacterial lineages [154]. In contrast, several amino acids found at the respective positions in NT-26 and *A. tumefaciens* FliM differ significantly. In particular, the presence of prolines at two of the crucial positions in NT-26 FliM suggests that the interaction between FliM and NT-26 CheY may be structurally different from the one observed in *E. coli*. Comparison of AgR and ExsF to *E. coli* CheY would therefore be inconclusive as different residues may be necessary for interacting with NT-26 FliM.



Protein	90	92	95	99	106	108	119	122
***E. coli* CheY**	**A**	**K**	**I**	**A**	**Y**	**V**	**K**	**K**
NT-26 CheY1	G	**R**	**V**	**A**	V	**A**	A	A
NT-26 CheY2	S	A	K	R	**W**	**V**	A	**R**
AgR	R	E	**L**	Y	**F**	**V**	A	D
ExsF	**A**	I/L	**V/L**	**A**	V/L	S/A	V/L	N/E

Figure 4.26: Comparison of AgR and ExsF with NT-26 and *E. coli* CheY proteins. A - multiple sequence alignment. Black arrowhead indicates the phosphorylatable aspartate, red - Y-T coupling Thr residue, blue arrowheads indicate residues interacting with the phosphoryl group. Empty red arrowhead signifies the variable position corresponding to *E. coli* CheY Tyr106. Residues involved in binding FliM peptide are shown as letters above the respective positions. ClustalX colouring scheme was applied. B - identity of residues corresponding to *E. coli* CheY amino acids interacting with FliM. Numbers denote residue positions in *E. coli* CheY.

Figure 4.27: Alignment of the N-terminal portion of *E. coli* FliM sequence and equivalent regions of NT-26 and *A. tumefaciens* FliM proteins. The residues involved in interaction with CheY are indicated by filled arrowheads (positions conserved across these sequences), empty arrowheads (semi-conserved positions) and crosses (variable positions). ClustalX colouring scheme was applied.

4.5.4 Sequence/structure predictions of AgR and ExsF and comparison to other receiver domains

As briefly discussed in Section 4.1.3, initial multiple sequence alignments against *E. coli* CheY and the receiver domain of ExsG revealed the presence of several unique features of both AgR and ExsF sequences. In order to verify if those unusual sequence elements are also found in homologous sequences, BLAST searches were performed using NT-26 AgR and ExsF sequences as queries. Interestingly, the closest ExsF homologue was 50% identical in terms of sequence while the closest homologues of AgR shared more than 80% sequence identity, suggesting that expansion of AgR-like RRs may have occurred more recently. Furthermore, ExsF homologues were confined primarily to α-proteobacteria, while close AgR homologues were identified in representatives of diverse bacterial phyla as well as in several archaeal species. It appeared that AgR-like sequences were highly conserved; for instance, NT-26 AgR shared over 40% sequence identity with homologous proteins from two archeons, *Halorhabdus utahensis* and *Halalkalicoccus jeotgali*. The residues identical across the sequences almost invariably involved several characteristic sequence motifs which are listed in Section 4.5.4.1.

The sequences of AgR and ExsF, as well as a number of their homologous sequences identified by BLAST searches, were compared to other single-domain RRs and receiver domains of several canonical RRs, in an attempt to identify homologues with known functions and verify the presence of the unique sequence motifs present in AgR and ExsF sequences. A multiple sequence alignment was generated using *Bacillus subtilis* Spo0F (UniProt entry P06628), a single-domain RR that participates in phosphorelay, and CheY proteins from various species. The sequences of several receiver domains contained within hybrid HKs (GacS, P48027; ArcB, P0AEC3; BarA, P0AEC5) were also included. The alignment (Fig. 4.28, for a large alignment see Appendix, Fig. B.4 and B.5) suggested that both AgR- and ExsF-like sequences constitute separate and

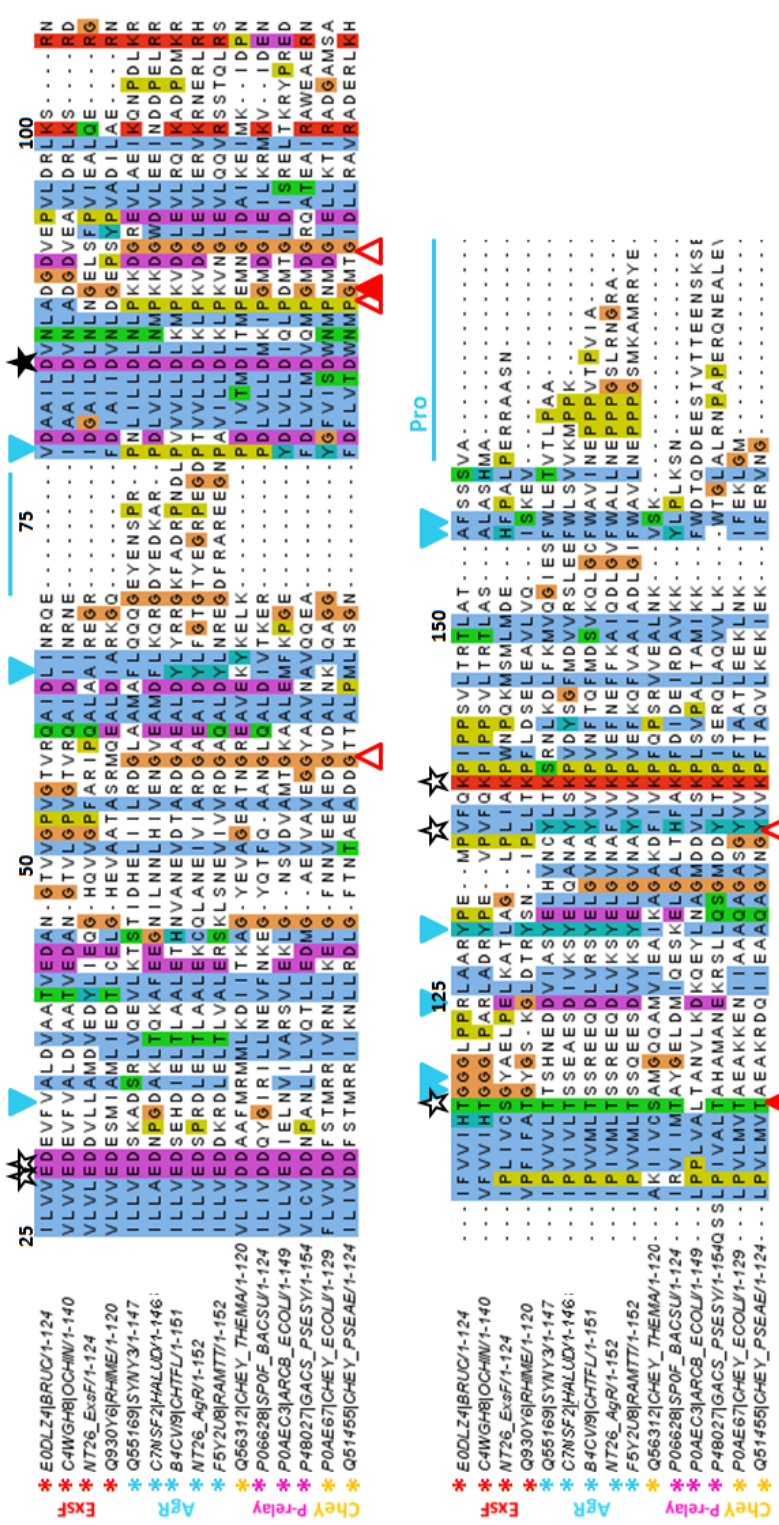


Figure 4.28: Multiple sequence alignment of AgR- and ExsF-like sequences against CheY and phosphorelay receiver domains. First 24 residues of the alignment are not shown. Characteristic sequence motifs of AgR are in cyan, of ExsF in red. Empty arrowheads indicate missing conserved amino acids. Black star denotes the phosphorylatable aspartate, empty stars - other active site and Y-T coupling residues. Cyan lines denote AgR sequence extensions, Pro - proline-rich region. BRUC - *Brucella* sp.; OCHIN - *Ochrobactrum intermedium*; RHIME - *Rhizobium meliloti*; SYNY3 - *Synechocystis* sp.; HALUD - *Halorhabdus utahensis*; CHTFL - *Chthoniobacter flavus*; RAMTT - *Ramlibacter tataouinensis*; THEMA - *Thermatoga maritima*; BACSU - *Bacillus subtilis*; ECOLI - *E. coli*; PSESY - *Pseudomonas syringae*; PSEAE - *Pseudomonas aeruginosa*. Codes represent UniProt entries. P-relay - phosphorelay. Note: the third ExsF motif (conserved Gly) has not been aligned properly here. ClustalX colouring scheme was applied.

distinct sub-classes of RRs. The extended α2/β3 loop and the C-terminal portion of AgR, as well as several highly conserved sequence motifs, were present in homologues from α-, β-, γ-, and δ-proteobacteria and many other bacterial phyla, including cyanobacteria, as well as archaea. The absence of two invariant glycines and one proline, and the presence of two conserved glycines at unique positions were the characteristic features across ExsF-like sequences, together with the replacement of the tyrosine involved in Y-T coupling with a valine or leucine.

There were a few sequence positions which could suggest that AgR and ExsF may be involved in phosphorelay. Both GacS and ArcB receiver domains contained a tryptophan close to the C terminus and an acidic residue at position 124 of the alignment, similarly to AgR, while the absence of the aromatic Y-T coupling residue characteristic of ExsF-like sequences is also observed in the case of Spo0F and ArcB. However, no other common motifs could be detected. Notably, similarities to ArcB and GacS receiver domains may be coincidental as these domains are not single-domain RRs - they are involved in intramolecular phosphorelay. In addition, as described in the section below, the "YW" or "FW" motif located towards the C-terminus of AgR-like sequences constitutes the centre of dimerisation patch. Given that these single-domain RRs exhibited unique features both in terms of sequence and structure, it was impossible to infer their likely function.

4.5.4.1 AgR subclass

Scanning against the CATH database with AgR sequence confirmed that the protein belongs to the CATH superfamily 3.40.50.2300 together with other single-domain RRs. It was classified into the functional family of "chemotaxis protein-like domain" proteins. In order to identify the closest structural homologue of AgR, the sequence was scanned against the Protein Data Bank using the BLAST search tool and against the CATH domain superfamily HMM profiles. In addition, it was submitted to PSIPRED webserver for fold analysis by pGenThreader. All three methods resulted in identification of several highly homologous sequences and structures. AgR was classified into CATH domain 1K66A00 represented by the structure of *Calothrix* PCC7601 RcpB (PDB ID 1K66 [57]), while based on pGen Threader analysis and BLASTp results the closest structural homologue of AgR was Rcp1 from *Synechocystis* sp. PCC6803 (PDB ID 1JLK and 1I3C [56]). RcpA, another AgR-like RR of *Calothrix*, was also shown in the analyses to be closely related to AgR.

Rcp proteins constitute single-domain RRs which are phosphorylated by cyanobacterial phytochromes (Cphs) [103, 57] and all seem to contain the unique highly conser-

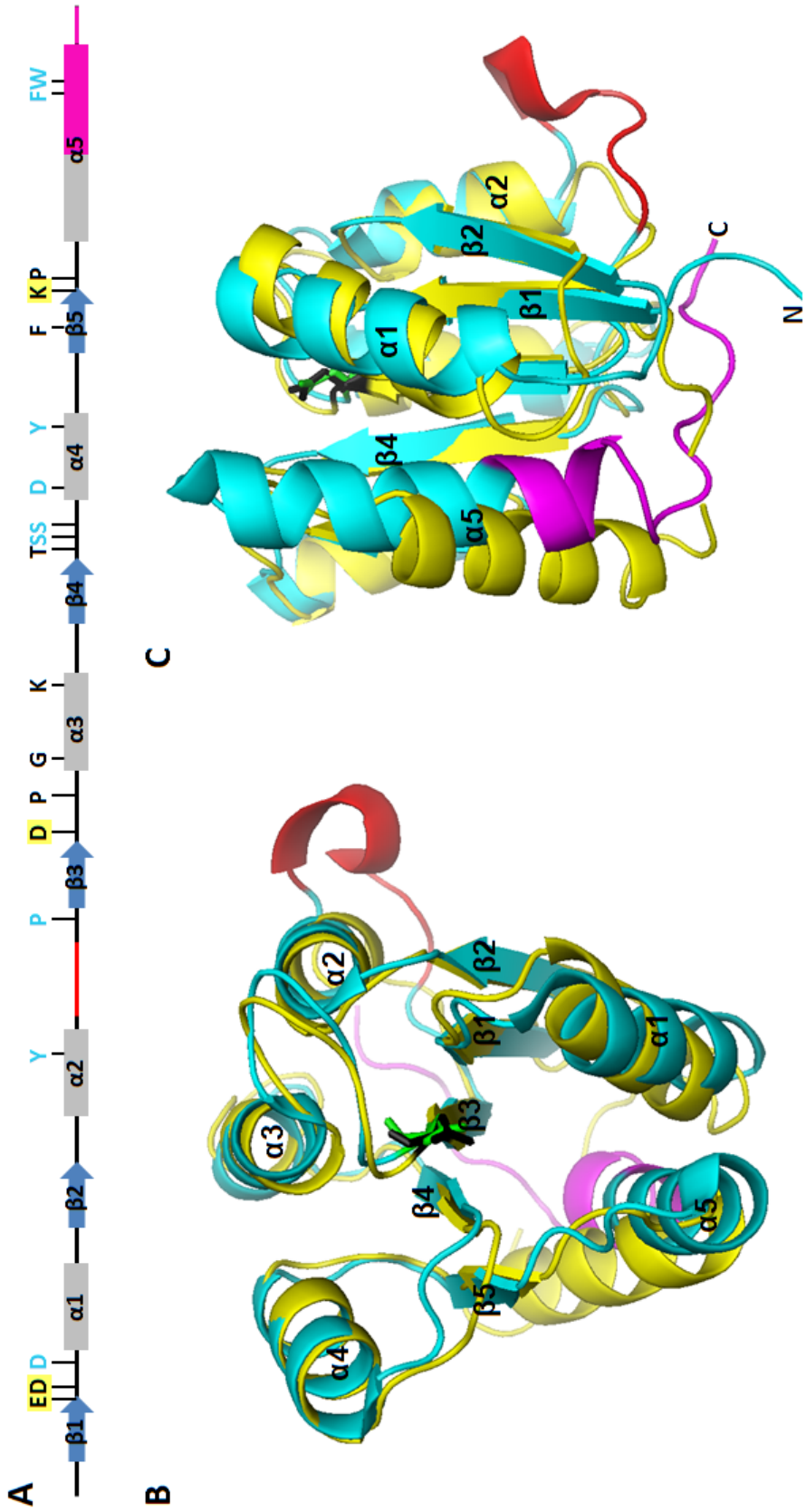


Figure 4.29: Characteristic sequence and structure motifs within AgR and comparison to CheY. A - schematic diagram (not to scale) showing predicted secondary structures of AgR and crucial amino acid positions. Helices are grey, strands are blue, residues conserved across all receiver domains are in black, active site residues are highlighted in yellow. The extended α2/β3 loop is in red, the extended C-terminal region in magenta, residues unique to AgR are in cyan. B and C - the structure of Rcp1 (PDB ID 1I3C, cyan) superimposed with CheY (3CHY [72], yellow), viewed from different angles; the unique AgR structural motifs are mapped and colour-coded as in (A). Active site aspartate of Rcp1 is shown as green sticks, CheY aspartate as black sticks.

ved sequence and structure motifs identified within AgR-like sequences (summary in Fig. 4.29A). The only notable exception is the proline-rich region close to the carboxyl end of AgR homologues; in Rcp proteins there is usually only one proline (see Fig. 4.28). Together with the α4/β5/α5 region, the extended C-terminal portion appears to form a part of highly conserved dimerisation interface, with the "FW" or "YW" motif (helix α5) located at the centre of the binding patch [57, 56]. Curiously, RcpA and RcpB were shown to exist as dimers irrespective of the phosphorylation state [57], in contrast to Rcp1 which exist as a monomer and dimerises upon phosphorylation [56]. As the α4/β5/α5 interface commonly mediates protein-protein interactions, e.g. FliM binding, its involvement in dimerisation and the presence of constitutive dimers raises the question about the localisation of the effector surface within AgR/Rcp proteins.

In order to visualise the structural differences between AgR/Rcp proteins and CheY, the canonical single-domain RR, the structure of Rcp1 was superimposed with CheY (Fig. 4.29B and C). All the secondary structure elements aligned well except for the extended α2/β3 loop and helix α5. The latter difference may be related to the fact that via the α4/β5/α5 region CheY interacts with FliM, while in AgR/Rcp proteins this surface is involved in homodimerisation. In the structure of Rcp1, as well as in the other Rcp structures, the extended loop sticks out and involves one helical turn. As it is present in all AgR homologues, and extrudes from the surface on the opposite side to the dimerisation interface, it may be functionally relevant.

Conservation of structural and sequence motifs within Rcp sequences was previously reported by Benda and co-workers, who suggested that Rcp proteins constitute a sub-class of CheY-like RRs [57]. The results presented herein further support the notion that single-domain RRs that belong to phytochrome signalling pathways form a considerable sub-class of RRs, characterised by unusually high sequence conservation across homologues in diverse bacterial and archaeal phyla. The presence of an extended loop extruding from the surface, and the fact that the α4/β5/α5 region is involved in homodimerisation, imply that AgR/Rcp RRs may exhibit a unique, and potentially novel, type of functionality.

4.5.4.2 ExsF subclass

The sequence of ExsF was scanned against the CATH database and, similarly to AgR, the protein was classified into the "chemotaxis protein-like domain" family. A BLAST search against the Protein Data Bank identified, however, only one protein, *Sinorhizobium (Rhizobium) meliloti* 0114, which gave a significant alignment score (86 bits). Sma0114 was also previously found as an ExsF homologue in the standard BLAST

search against the non-redundant sequence database, and shares 36% identity with ExsF. No CATH domain was assigned to ExsF.

Analogously to ExsF, Sma0114 is a single domain response regulator encoded next to a HWE HK, and lacks several crucial RR sequence motifs - most importantly, the tyrosine involved in Y-T coupling is replaced by a leucine [155]. The absence of an aromatic residue has only been observed in the case of receiver domains involved in phosphorelay (see Fig. 4.28). The other Y-T coupling residue, threonine, appears to be a part of a "FAT GUY" sequence motif (PFxFATGY) [155]; according the the multiple sequence alignment assembled (see Fig. 4.28 and Appendix, Fig. B.4 and B.5), the glycine residue of that motif is highly conserved across all ExsF-like sequences. The NMR assignment of Sma0114 revealed that, in agreement with the secondary structure prediction from PSIPRED, the protein lacks the fourth helix. In NT-26 ExsF the predicted helix α4 was only 4 residues long (see Fig. 4.3) and the PSIPRED prediction was, in fact, associated with a low probability (50%). Sma0114 is a monomer and ExsF seems to exist as a monomer as well (see Chapter 5, Section 5.2.2), but it remains unknown whether phosphorylation induces dimerisation as in the case of Rcp1 [56].

Sheftic and co-workers suggested that the interaction between HWE HKs and their cognate RRs may involve a mechanism distinct to that observed in the case of canonical proteins. The low level of phosphotransfer between BphP1-HK and ExsF observed in the transphosphorylation assay (see Fig. 4.19) demonstrates that ExsF-type RRs do not differ from canonical RRs in terms of the phosphotransfer mechanism or the mode of interaction with the cognate HK. However, they may exhibit differences in the context of activation and the form of protein-protein interactions they engage in. Given that the conserved tyrosine residue and helix α4 are absent from ExsF-like RRs, activation of the receiver domain may involve a mechanism different from Y-T coupling. The absence of helix α4 from the region commonly mediating protein-protein interactions also suggests that these RRs may regulate unusual downstream protein targets.

Several attempts were made in the course of this project to crystallise ExsF. A range of protein concentrations and over 700 crystallisation conditions were tested but no crystals were observed. Crystallisation attempts in the presence of magnesium chloride or beryllofluoride and magnesium chloride were also unsuccessful.

4.6 Summary

The results described in this chapter show that the four proteins belonging to the BphP1 gene cluster form a branched two-component signalling pathway, and BphP1 exhibits

a phosphorylation preference for ExsG over the other cognate RR, AgR. ExsG was demonstrated to possess both RR and HK activity; the relationship between the two domains and their activities is further discussed in Chapter 6. In addition, the SEC profile of ExsG indicated that the protein may form a higher-order structure, and this is further explored in Chapter 5.

Despite various attempts and approaches, determination of the response triggered by the BphP1-initiated signalling cascade was unsuccessful. Based on the temporal stability of the phosphoresidues and the apparent absence of phosphatase activity, it was concluded that the signalling cascade investigated herein most probably affects gene expression rather than evokes a rapid response requiring an instantaneous adaptation. Given that AgR and ExsF most likely constitute the final steps within the pathway, they were tested for the ability to relay the signal onto the Hpt domain protein and were compared against other, well-characterised single-domain RRs. However, the results were inconclusive primarily due to the presence of a number of unique sequence and structure motifs within the two proteins. It appears that despite the overall similarity to canonical receiver domains, ExsF- and AgR-like proteins belong to distinct sub-classes of single-domain RRs. The divergence of ExsF most likely reflects the different mode of interaction with HWE HK core. The involvement of the α4/β5/α5 interface in homodimerisation of AgR/Rcp proteins, and the presence of the protuberant loop, suggest that these proteins may exhibit unique functionality.

Chapter 5

Biophysical and structural characterisation of ExsG

ExsG is an unusual hybrid HK which contains an uncommon HWE HK core with a receiver domain appended at the amino terminus. As described in Chapter 4, ExsG exhibits both HK and RR activity and, based on the size-exclusion chromatography profiles, it seemed that it may form a large complex. At the time when this thesis was written no structural information was available on proteins such as ExsG. Therefore several biochemical and biophysical techniques were applied to characterise this protein and determine its three-dimensional structure.

5.1 Determination of the identity of ExsG-associated low-intensity protein bands

As mentioned in Chapter 4, two contaminating low-intensity protein bands were observed above the main ExsG protein band on SDS-PAGE gels after the purification tag was cleaved off. They could not be removed by SEC and appeared to associate with all ExsG site-specific variants, except for H151N. Furthermore, in the case of each ExsG protein product the main protein band migrated further on SDS-PAGE gel than would be expected for a 37.6 kDa protein, with the apparent molecular weight of ∼30-34 kDa. To determine the identity of the proteins contained within each of the three bands, the proteins were extracted from the gel, subjected to tryptic digestion and analysed using liquid chromatography- tandem mass spectrometry (LC-tandem MS). The bands used in the analysis were from both wild-type ExsG and the D62N protein variant; the results were largely identical, hence only the results from WT ExsG analysis are shown

below.

Even though not all peptides that should have been generated by the proteolytic digestion were detected in the analysis of either protein, all the identified protein fragments from each gel band matched the sequence of ExsG (see Figures in Appendix C for lists of peptides and Fig. 5.1 for sequence coverage map), demonstrating that each of the three SDS-PAGE bands represents a form of ExsG. Peptides obtained from the top band (colour-coded blue in Fig. 5.1) covered 62% of ExsG sequence, while those from the middle (red) and bottom (green) bands incorporated 78% of the sequence. The analysis revealed that the first residue (glycine) of the protein product contained within the top band was acetylated, while the middle band contained protein that was phosphorylated upon Ser163. No covalent modifications were detected in peptides derived from the main protein band. Unfortunately, the peptide containing the phosphorylatable His residue was not recovered and detected in the analysis. As H151N protein product appeared as a single band on SDS-PAGE gels, it seemed that the active site histidine may in some way affect the efficiency of serine phosphorylation and N-terminal acetylation; it could also be itself subject to post-translational modifications.

5.2 Determination of ExsG oligomeric state and assembly pathway

Initially, size-exclusion chromatography using Superose 6 was primarily a purification step. However, even though this column exhibited poor resolution of the proteins, it was evident that WT ExsG and its five site-specific variants eluted significantly earlier than all the other proteins belonging to the BphP1 signalling cascade, suggesting that ExsG forms a higher order structure (see Chapter 4). Further evidence for ExsG oligomerisation was obtained from matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-ToF MS) (courtesy of Sibylle Heidelberger, School of Pharmacy). This technique was applied in order to verify the monomeric mass of ExsG, as the apparent molecular mass of ExsG based on SDS-PAGE analysis ($\sim$32 - 34 kDa) was lower than the calculated one (37.6 kDa). Curiously, despite the fact that MALDI-ToF MS normally disrupts non-covalent interactions, dimeric, trimeric, and even tetrameric states of ExsG were detected in the acquired spectra (data not shown).

A more extensive comparison of the elution profiles, together with estimation of molecular weights, is presented herein. To identify the region of ExsG responsible for the formation of higher order structure, the HK core was cloned, expressed and purified

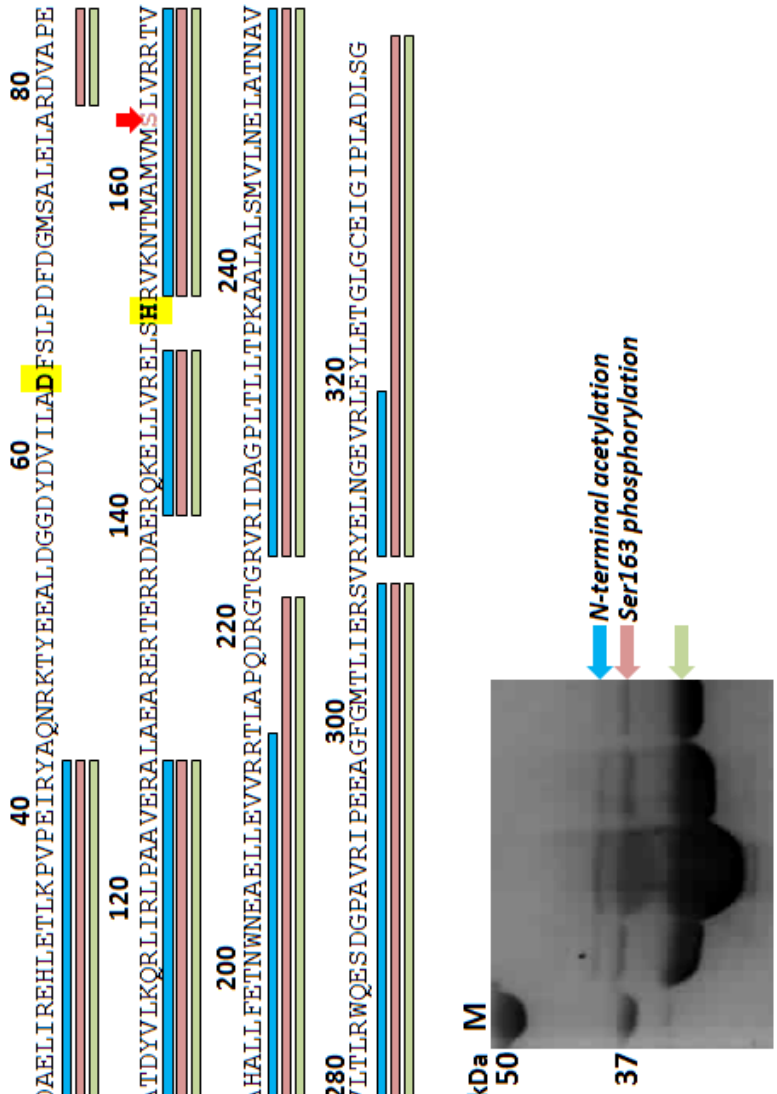


Figure 5.1: LC-tandem MS analysis of ExsG protein bands. Top: map of ExsG sequence coverage by peptides detected for each protein band; blue - top band, red - middle band, green - bottom (main) band. Phosphorylatable residues are highlighted in yellow and the two covalently modified residues are indicated with red arrows. Data acquired and analysed by Adam Cryar. Bottom: SDS-PAGE gel showing the respective protein bands and the associated post-translational modifications.

as a separate protein product. The molecular weights of ExsG-REC and ExsG-HK were estimated based on data obtained from preparative and analytical SEC as well as native mass spectrometry. In order to accurately determine the native state of ExsG, both analytical ultracentrifugation and native mass spectrometry were applied.

5.2.1 Expression, purification and activity test of stand-alone HK core of ExsG (ExsG-HK)

The kinase core portion of ExsG was initially sub-cloned (see Fig. 5.2A for the expression construct), expressed and purified to establish if it retains activity without the receiver domain. The relationship between the activity of the receiver domain and the kinase core is described in Chapter 6. Here ExsG-HK, along with ExsG-REC, serve to demonstrate which of the two functional regions of ExsG is responsible for oligomerisation.

The results of the two stages of ExsG-HK purification are shown in Figure 5.2B and C. The protein (27 kDa with purification tag, 23 kDa without) was expressed and purified in the same way as the WT ExsG, but the yield from NiNTA purification was considerably lower (∼1-1.5 mg per 1 g of cells). After removing the purification tag, the protein was subjected to SEC using Superose 6. The elution peak was not symmetrical - it seemed that an additional small peak (indicated by green arrow in Fig. 5.2C) was partially overlapping with the main peak. As the eluted protein appeared pure in SDS-PAGE analysis, this suggested that the protein may exist in different oligomeric states. The two covalently modified protein forms, which were observed in SDS-PAGE analysis of most full-length ExsG protein products (see previous section), were not detected. Notably, ExsG-HK eluted earlier (15 - 15.5 ml) than BphP1-HK (16.5 - 17 ml, 27 kDa), which implied that both WT ExsG and ExsG-HK might form larger complexes than BphP1-HK.

To confirm that ExsG-HK protein product is active, an autokinase assay was performed. The protein was incubated with the "hot" reaction buffer and samples were taken at several time points over 4 hours. From the resulting autoradiograph it was evident that ExsG-HK exhibits autokinase activity (Fig. 5.3A), although the autophosphorylation rate (∼4.4 PSL min^{-1} $pmol^{-1}$) was 5-fold lower than that of WT ExsG (22.6 PSL min^{-1} $pmol^{-1}$) (Fig. 5.3B). As described later in Section 5.4, a stretch of 22 amino acids, encompassing a part of the interdomain linker and ending with the phosphorylatable histidine, exhibits a pattern suggesting that it may form a coiled-coil structure. Even though the entire region was incorporated into the ExsG-HK construct, without

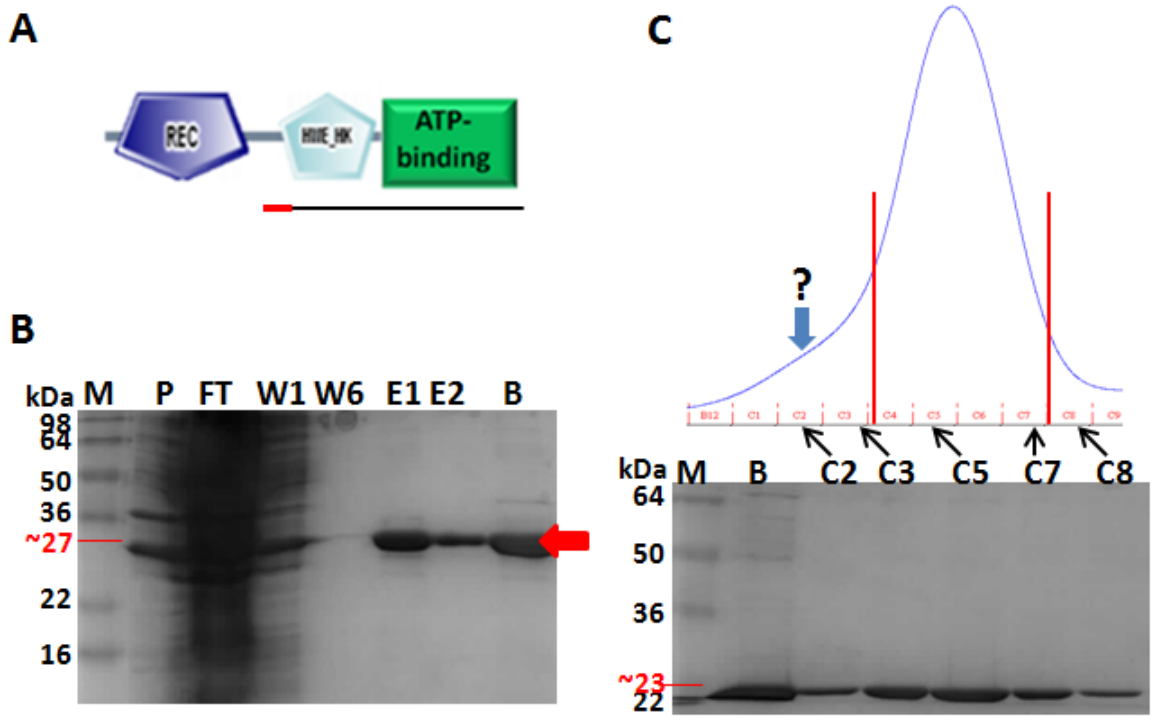


Figure 5.2: Purification of ExsG-HK protein product. A - schematic representation of ExsG-HK expression construct. Red line indicates 18 residues of the interdomain linker that were incorporated into the construct. B - SDS-PAGE analysis of NiNTA purification. M - marker, P - pellet (insoluble fraction), W - wash, E - elution, B - beads. Red arrow indicates the protein with the N-terminal purification tag (∼27 kDa). C - size-exclusion chromatography using Superose 6: partial elution profile (top) and SDS-PAGE analysis (bottom). The blue arrow points at a small peak that overlaps with the main protein peak and red lines denote the range of pooled fractions. SDS-PAGE analysis shows corresponding elution fractions and the position of the protein (without purification tag) is indicated in red. M - marker, B - ExsG-HK before SEC.

the remaining N-terminal portion of the protein the coiled-coil may not form properly. This may affect the autokinase activity as the helical interdomain linker leads directly into the helix containing the active site histidine.

5.2.2 Comparison of size-exclusion chromatography profiles and molecular mass estimation of ExsG protein products

The SEC profiles of the three ExsG protein products - WT ExsG, ExsG-HK and ExsG-REC - were compared against the profiles of BphP1-HK, ExsF and AgR to establish the relative molecular masses, and verify if ExsG protein products have a propensity to form a complex greater than a dimer (Fig. 5.4). Both WT ExsG and ExsG-HK eluted

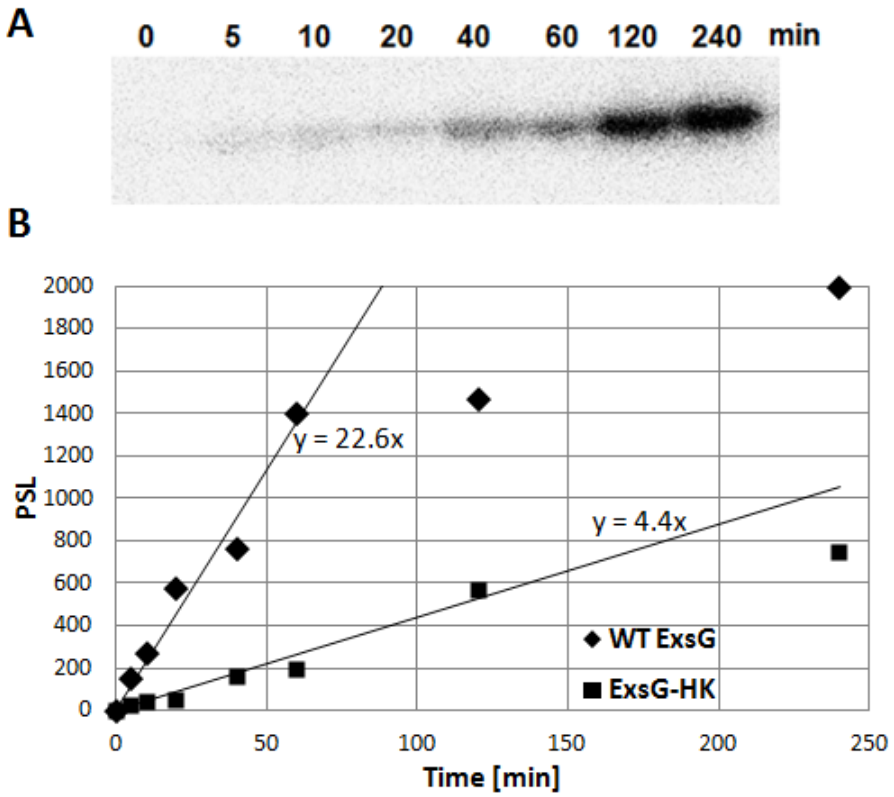


Figure 5.3: Autokinase assay of ExsG-HK and comparison of WT ExsG and ExsG-HK autophosphorylation rates. A - autoradiograph showing the results of autokinase assay. B - plots of the PSL signal versus time for WT ExsG and ExsG-HK. Trendlines were based on the linear ranges - up to 120 min for ExsG-HK and up to 60 min for WT ExsG.

earlier from the Superose 6 column than BphP1-HK, despite the fact that BphP1-HK is larger in terms of monomeric mass than ExsG-HK. This would suggest that ExsG-HK and WT ExsG form higher-order structures, under the assumption that BphP1-HK, comprising a canonical HK core, is a homodimer.

On the other hand, ExsG-REC eluted later than BphP1-HK and the elution peak overlapped with that of ExsF and AgR, implying that the native states of the single-domain RRs are of similar molecular mass. The estimated molecular mass based on the calibration curve (Fig. 5.4B) indicated that ExsG-HK (288 kDa) is almost twice as large as BphP1-HK (151 kDa), and about 3.5-times greater than the three single-domain RRs (79 kDa), while WT ExsG, with an estimated molecular weight of 480 kDa, is 6-times greater than the RRs and about 3-times greater than BphP1-HK. Given the poor resolution achieved with Superose 6, these apparent sizes are unlikely to reflect actual molecular masses, but the size relationships may still apply.

Canonical HKs characterised so far have been shown to be consistently homodimeric which also applies to HK cores found in bacteriophytochromes - BphPs from *Deinococ-*

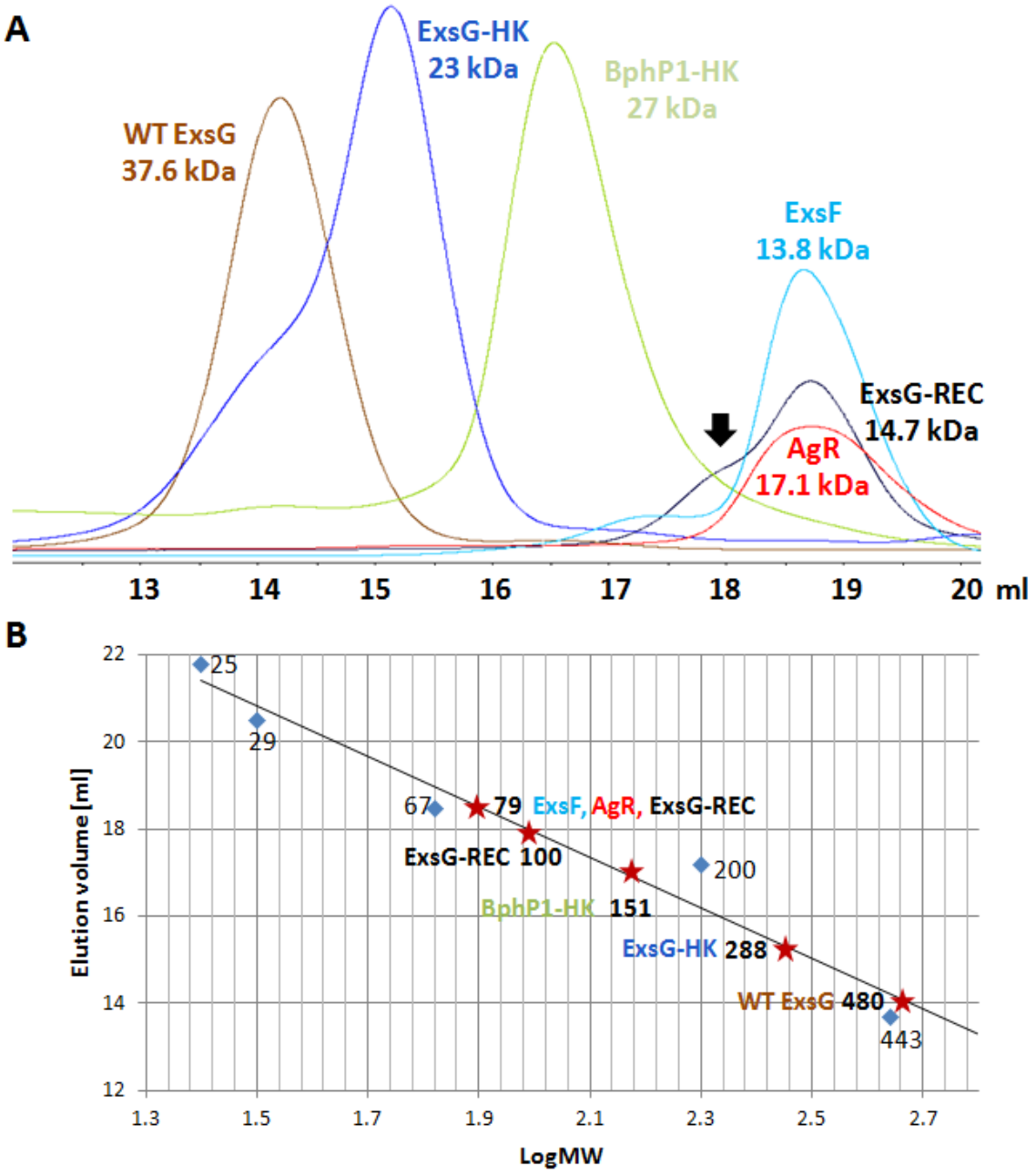


Figure 5.4: Comparison of SEC elution profiles of AgR, ExsF, BphP1-HK and ExsG protein products from Superose 6. A - overlay of SEC profiles. Black arrow indicates the higher oligomeric state formed by ExsG-REC. B - calibration curve with marked apparent molecular masses of all protein products. Molecular mass (weight) markers (blue diamonds): chymotrypsin (25 kDa), carbonic anhydrase (29 kDa), albumin (67 kDa), amylase (200 kDa) and apoferritin (443 kDa). Red stars denote the positions of the relevant calculated molecular masses of the protein products.

cus radiodurans and *Pseudomonas aeruginosa* have been demonstrated to form dimers [109, 105]. Hence it seemed likely that BphP1-HK, which constitutes a canonical HK core, is a homodimer. To verify the oligomeric state of BphP1-HK, the protein was subjected to SEC using an analytical Superdex 75 column. The obtained profile (colour-coded blue in Fig. 5.5) showed that the protein did not elute in the void volume, and the estimated molecular mass (71 kDa) could correspond to a dimeric state (54 kDa). Larger apparent molecular mass may be due to the likely non-globular shape of the HK core (see Introduction, Fig. 1.2 and Section 1.2.1.1). BphP1-HK therefore appeared to be a dimer, at least under the applied conditions, and this provided further evidence that WT ExsG and ExsG-HK must indeed form higher-order structures as they eluted significantly earlier in SEC.

To determine if ExsG-REC promotes oligomerisation, its molecular mass was also estimated based on the elution profile from the analytical column (red in Fig. 5.5). With the apparent molecular mass of 25 kDa, the protein was likely to exist as a monomer or a dimer. As ExsG-REC was susceptible to degradation and precipitation, attempts were made to accurately determine the molecular mass of ExsF, which was more stable and soluble. The elution profile from analytical SEC revealed that ExsF was also a monomer or a dimer, with an estimated molecular mass of 22 kDa (green in Fig. 5.5). By means of electrospray ionisation mass spectrometry (ESI MS) it was established that ExsF exists as a monomer in the gaseous phase (13.7 kDa, Fig. 5.6), which suggests that it could be a monomer in solution as well. Therefore ExsG-REC is also most likely a monomer, or at most a dimer, indicating that it is not sufficient for oligomerisation. In conclusion, ExsG oligomerises via the HWE HK core, while the receiver domain does not contribute to the formation of the higher-order structure but may exhibit a dimerisation propensity as indicated by the presence of an additional peak on the elution profile (Fig. 5.4).

5.2.3 Analytical ultracentrifugation (AUC) analysis of wild-type ExsG

WT ExsG was subjected to two different AUC experiments in order to determine the molecular mass and to gain basic information about the shape of this macromolecule.

5.2.3.1 Sedimentation velocity experiment

This type of experiment allows determination of sample homogeneity as well as estimation of the molecular mass and frictional ratio, which is a parameter that describes the

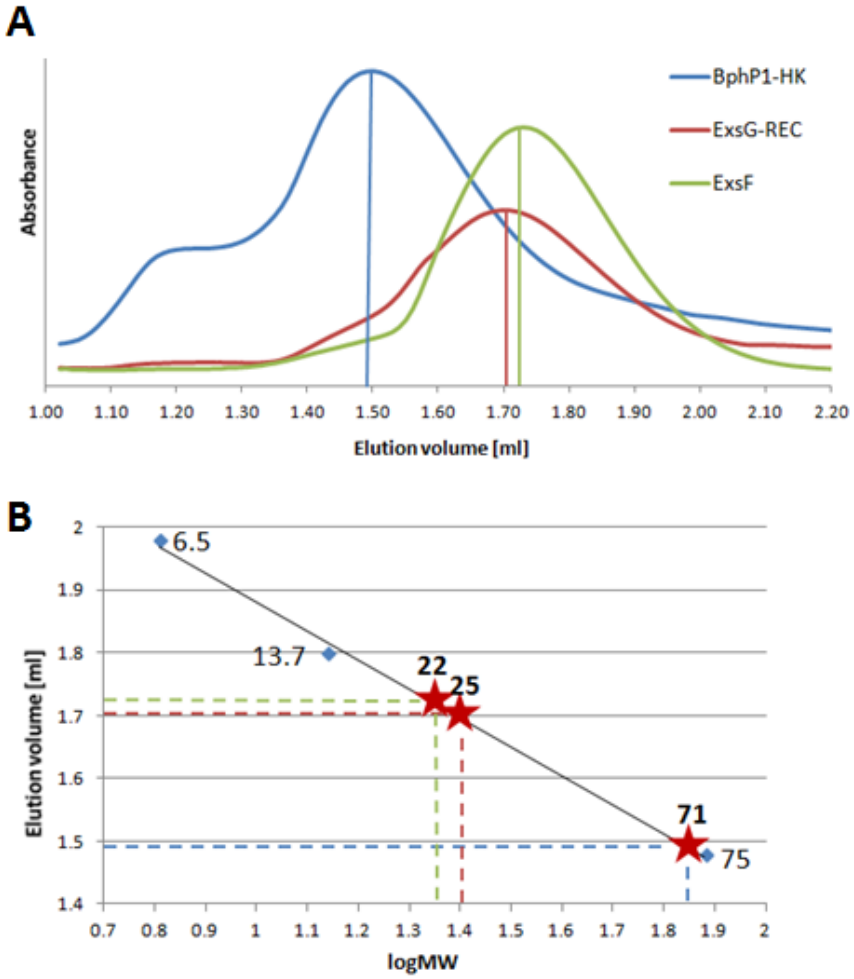


Figure 5.5: Elution profiles of BphP1-HK, ExsG-REC and ExsF from analytical SEC and molecular mass estimation. A - partial elution profiles; vertical lines show the peak positions used to determine the elution volumes. B - calibration curve with annotated molecular mass estimates (red stars) of BphP1-HK (71 kDa), ExsG-REC (25 kDa) and ExsF (22 kDa). Markers used for calibration are denoted by blue diamonds: conalbumin (75 kDa), ribonuclease A (13.7 kDa), aprotinin (6.5 kDa).

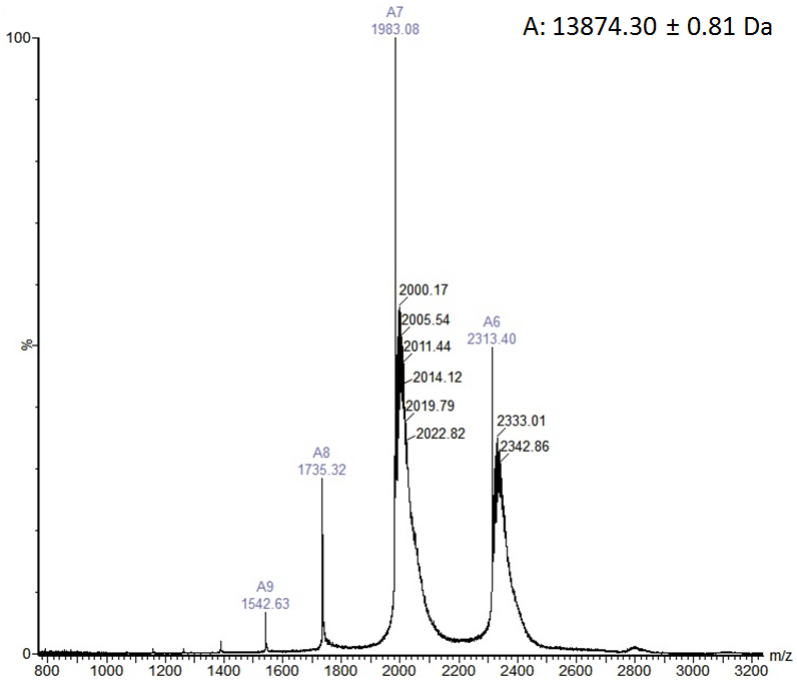


Figure 5.6: Native mass spectrum of ExsF. Data acquired and analysed by Jun Yan.

overall shape of the protein (globular vs ellipsoid). This information is obtained from a direct fit of a number of parameters to the experimental data. Notably, the signal is expressed in terms of "fringes" as the data was acquired using the interference optical system, which produces signal in the form of horizontal fringes. Changes in concentration boundaries along the sample cell affect the refractive index difference between the sample and reference solutions, leading to vertical displacement of these fringes.

Protein solutions at 7 different concentrations (5 - 55 μM) of WT ExsG were initially subjected to centrifugation at 50000 rpm. However, after 20 minutes most of the protein was found to have sedimented to the bottom of the cell, and therefore a new run with freshly prepared protein samples was set up with the speed reduced to 40000 rpm. Following the experiment, the protein was examined using SDS-PAGE to assess the effect of centrifugation of protein integrity.

Data analysis using the programme SEDFIT [124, 156] allows some parameters, such as the position of the cell meniscus and frictional ratio, to be floated in the fitting process. Thus individual parameters, or a combination of them, can be adjusted in order to reduce the root mean square deviation (RMSD) of the fit. The output of the analysis is a sedimentation coefficient distribution, which can be converted into a molecular mass distribution. Initial analysis of ExsG data was based on a frictional ratio 1.2 - 1.4, which describes a globular protein, but no good fits to the data could be

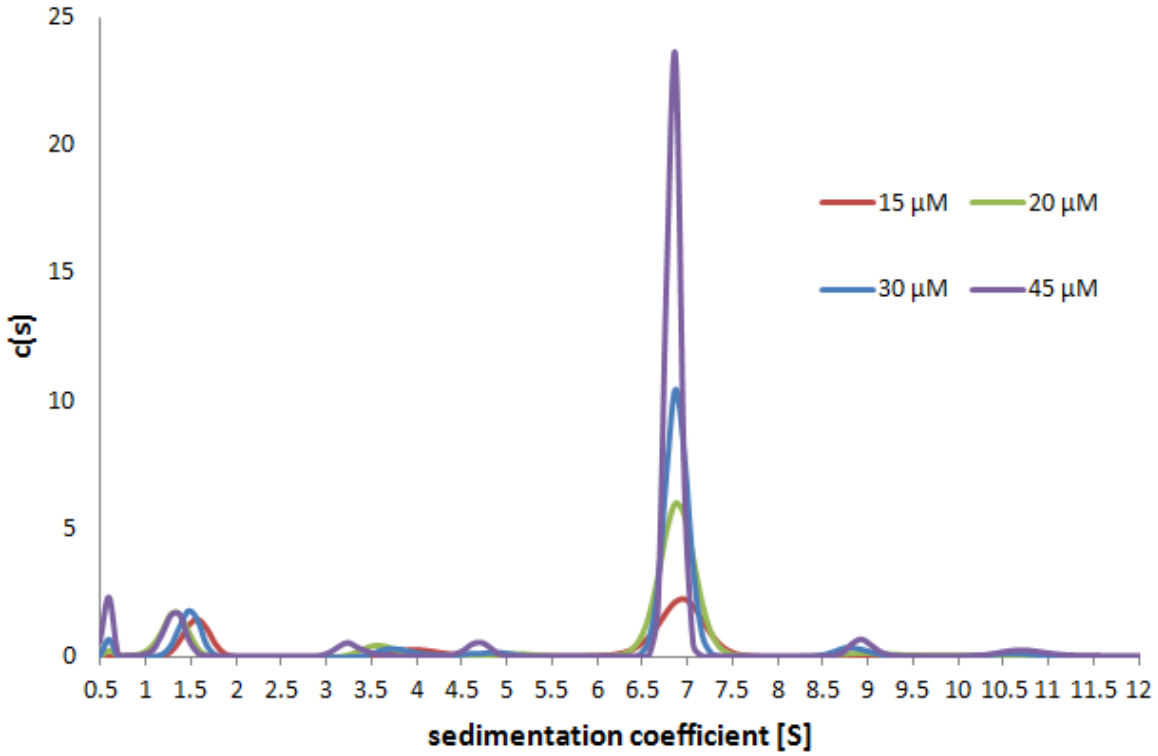


Figure 5.7: AUC sedimentation velocity analysis of ExsG. Sedimentation coefficient distribution, *c(s)*, was produced by SEDFIT for each ExsG sample; here the distributions for 4 different ExsG concentrations are shown. ExsG constitutes the main sedimenting peak at 6.8 Svedberg units (S). Data acquired by Jayesh Gor.

obtained except for the cell containing the lowest protein concentration, where RMSD was below 0.01 fringes. However, signal to noise ratio for the 0.2 mg/ml sample was very low and the reliability of the fit was dubious.

In the next round of analysis, the frictional coefficient was allowed to float during the fitting process, and it settled around 1.75, which corresponds to an elongated (ellipsoid) shape of the macromolecule. Further improvement was achieved by floating both the frictional ratio and the meniscus position. For each ExsG concentration a similar frictional ratio (1.7 - 1.95) and sedimentation coefficient distribution (Fig. 5.7) were obtained, and the molecular mass estimates for the main sedimenting species (∼6.8 S) were in the range 230 - 250 kDa. The RMSD values for the fits were normally below 0.01 fringes, except the data obtained from the two highest concentrations of ExsG, where signal to noise ratio was very high and the RMSD was between 0.01 and 0.05 fringes.

It therefore appeared that ExsG most likely forms an elongated hexamer (226 kDa) or heptamer (263 kDa). Sedimentation coefficient distribution obtained from different ExsG concentrations indicated the presence of one main sedimenting species, suggesting

that the protein does not adopt additional oligomeric states.

5.2.3.2 Sedimentation equilibrium experiment

In sedimentation equilibrium the protein solution is centrifuged at a low rotor speed for a long period of time to allow the formation of a stable concentration gradient, when sedimentation is at an equilibrium with diffusion of macromolecules down the concentration gradient. The shape of the equilibrium gradient is independent of the shape of the macromolecule, which permits an accurate determination of the molecular mass.

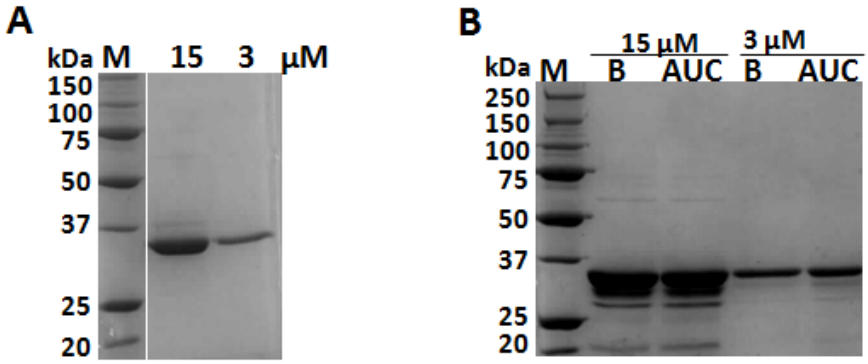


Figure 5.8: SDS-PAGE analysis of ExsG before and after AUC sedimentation equilibrium experiment. A - ExsG at two concentrations prior to loading into the sample cells. B - ExsG samples after the experiment; AUC - samples extracted from centrifugal cells, B - samples left at room temperature on the bench for the equivalent period of time.

Samples of ExsG at 6 concentrations of ExsG (1 - 40 µM) were subjected to 4 rotor speeds (7, 11, 16 and 24 krpm), and for each speed the system was allowed 30 hours to equilibrate. As the experiment lasted over 5 days and was performed at 20 °C, the protein samples extracted from the centrifugal cells after the experiment were analysed by SDS-PAGE and compared to samples prior to centrifugation to evaluate protein integrity at the end state of the experiment. In addition, a comparison was made against protein samples left for an equivalent length of time on the bench at room temperature. As can be seen in Figure 5.8, protein degradation was evident both in the samples subjected to centrifugation and those incubated on the bench, with similar proportion of degraded protein in each case. However, as was later seen in data analysis, this partial time-dependent fragmentation did not seem to have a significant impact on the data acquired towards the end of the experiment. Absorbance and interference scans taken every 3 hours were used in SEDFIT to determine if equilibrium was reached, and the equilibrium data was exported to the programme SEDPHAT. Similarly to SEDFIT,

SEDPHAT searches for the best fit to the experimental data, but it also offers the option to perform global fits, on several sets of data at once. A global chi-squared value is thus obtained, which for an ideal fit would be equal to 1. Both individual and global fits were attempted, including a combination of absorbance and interference data. Furthermore, where possible, multispeed files were assembled - such files contain equilibrium gradient data for several speeds, recorded for each protein sample.

It was impossible to obtain a high-quality and consistent fit to each data set with the molecular weight as a floating parameter, most likely due to partial degradation over time and owing to the fact that the samples were not subjected to extensive dialysis prior to the experiment, which could have resulted in artefacts. The molecular weight was therefore constrained to 226 kDa (hexamer), based on the sedimentation velocity results, while protein concentration, cell bottom and meniscus positions, as well as extinction coefficient, were allowed to float. The parameters were floated either one at a time or, if the fit could not be improved significantly, they were floated in pairs or all at once. This approach yielded relatively good fits (RMSD < 0.05) for most data sets, particularly those acquired from low (< 20 μM) protein concentration samples, although some fits were rejected as occasionally the optimised parameters were unrealistic (e.g. the protein concentration loaded). An example of SEDPHAT output for a fit to a single multi-speed absorbance file is shown in Fig. 5.9. Fitting with the weight constrained to 263 kDa (heptamer) was not attempted. For global runs, involving 2 or more data files, the lowest chi-squared value obtained was ∼2.2, which together with low local RMSD values (< 0.01) indicated a relatively good fit. However, for samples containing a protein concentration greater than 20 μM, the chi-squared value normally exceeded 4.

Even though only some of the fits to sedimentation equilibrium data were satisfactory, it still seemed that the molecular weight of ExsG is above 200 kDa, suggesting that the protein forms at least a homohexamer. At higher protein concentrations the quality of the fits decreased possibly due to protein aggregation, or because of the high signal intensity.

5.2.4 ExsG studies by electrospray ionisation mass spectrometry (ESI MS)

As the results from both AUC experiments suggested that ExsG forms at least a hexamer, but the molecular mass could not be accurately determined from the acquired data, the native state of the protein was determined in the gaseous phase using ESI

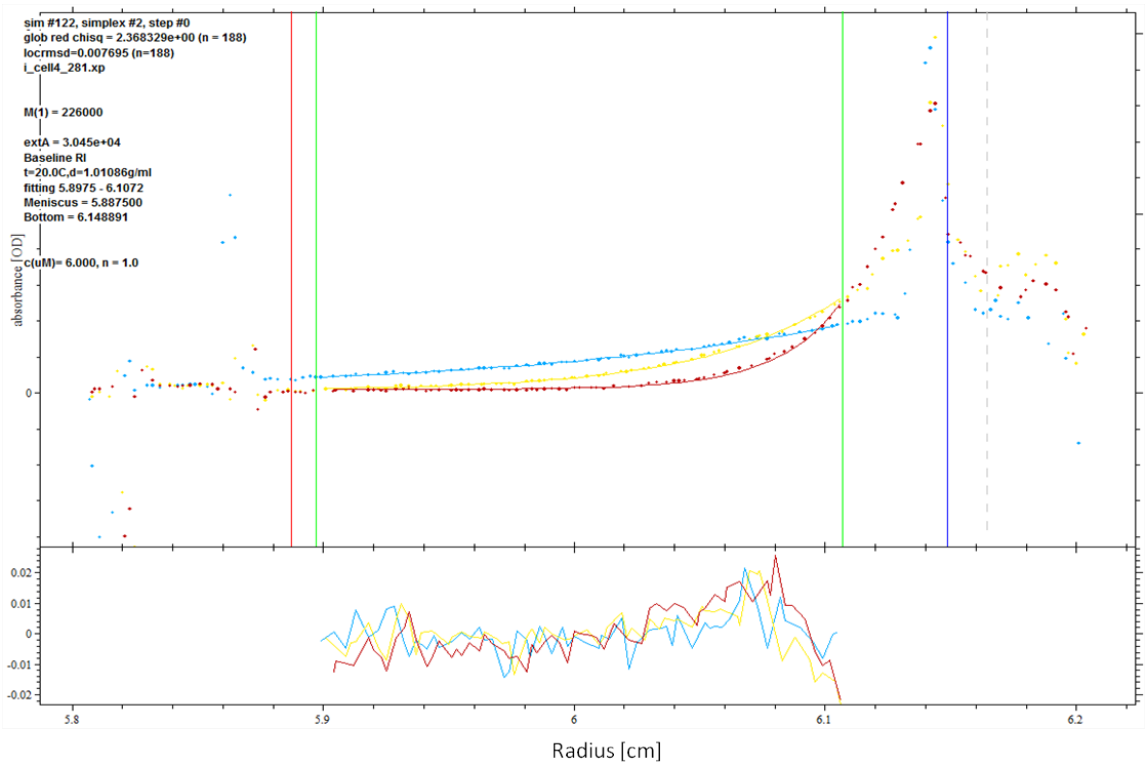


Figure 5.9: Example result from AUC sedimentation equilibrium data analysis. Multispeed file with absorbance data for 7000 rpm (blue dots), 16000 rpm (yellow) and 24000 rpm (red). The lines of corresponding colours indicate the fitted curves. Green vertical lines denote the range of fitting, red - meniscus and blue - cell bottom position. Parameters listed: glob red chisq - global reduced chi-squared, locrmsd - local RMSD, M (1) - molecular weight (constrained), extA - extinction coefficient (floated), t - temperature, d - buffer density, c(uM) - protein concentration (constrained). The panel below shows the RMSD distribution (deviation of fitted curve from experimental data) in absorbance units. Data acquired by Jayesh Gor.

MS. In addition, an attempt was made to establish the native state of ExsG-HK protein product as, based on the observations from SEC, it seemed that the HK core of ExsG is sufficient for oligomerisation. Full-length ExsG protein products were soluble in ammonium acetate solution, normally used in ESI MS, but ExsG-HK precipitated out during extensive buffer exchanges into different concentrations of ammonium acetate (30, 150 and 300 mM), and was not detected later in the MS analysis. SDS-PAGE analysis of WT ExsG showed that the covalently modified forms were still present after the buffer exchange (indicated by arrows in Fig. 5.10A), and the protein retained the ability to autophosphorylate as verified by an autokinase assay (Fig. 5.3B).

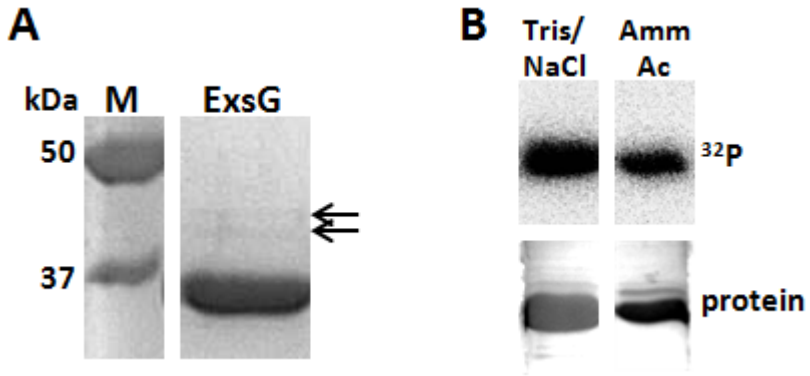


Figure 5.10: Autokinase assay and SDS-PAGE analysis of WT ExsG in ammonium acetate buffer. A - a magnified image of SDS-PAGE gel showing the protein band of ExsG in ammonium acetate. Arrows point at the bands containing covalently modified forms of ExsG. B - autokinase assay of ExsG in standard "hot" reaction buffer (Tris/NaCl) and "hot" ammonium acetate buffer.

5.2.4.1 Determination of the native mass of WT ExsG

Both WT ExsG and the H151N variant were extensively buffer exchanged into 150 mM ammonium acetate containing 100 μM DTT and sprayed into the MS chamber. The initially observed protein aggregation, which caused discontinuous spraying, was eliminated by raising the pH to 8.2. The spectra obtained for both ExsG protein products indicated the molecular weight of 226 kDa, which confirmed that ExsG is a homohexamer (Fig. 5.11A). Dimers, tetramers and even dodecamers were also detected, although at a significantly lower intensity.

Interestingly, in the spectra acquired for both WT ExsG and H151N variant, the peaks frequently exhibited a "shoulder", which suggested that an additional species (228 kDa) is associated with the WT protein; its abundance differed in a prep-dependent manner. Increasing the collision energy in the MS chamber allowed a better resolution of this low-intensity peak (Fig. 5.11B). The species could correspond to a covalently modified form of ExsG (see Section 5.1), but neither N-terminal acetylation nor Ser163 phosphorylation were likely to account for the observed 2 kDa difference. Analysis of these protein species using tandem-MS was impossible as it was not present at a sufficient concentration.

5.2.4.2 Elucidation of ExsG hexamer assembly pathway

Some oligomeric states of ExsG (trimers and pentamers) were not observed in the mass spectra, while dimers were detected at a higher intensity than tetramers and dodecamers. This suggested that dimers may constitute the basic structural unit for the

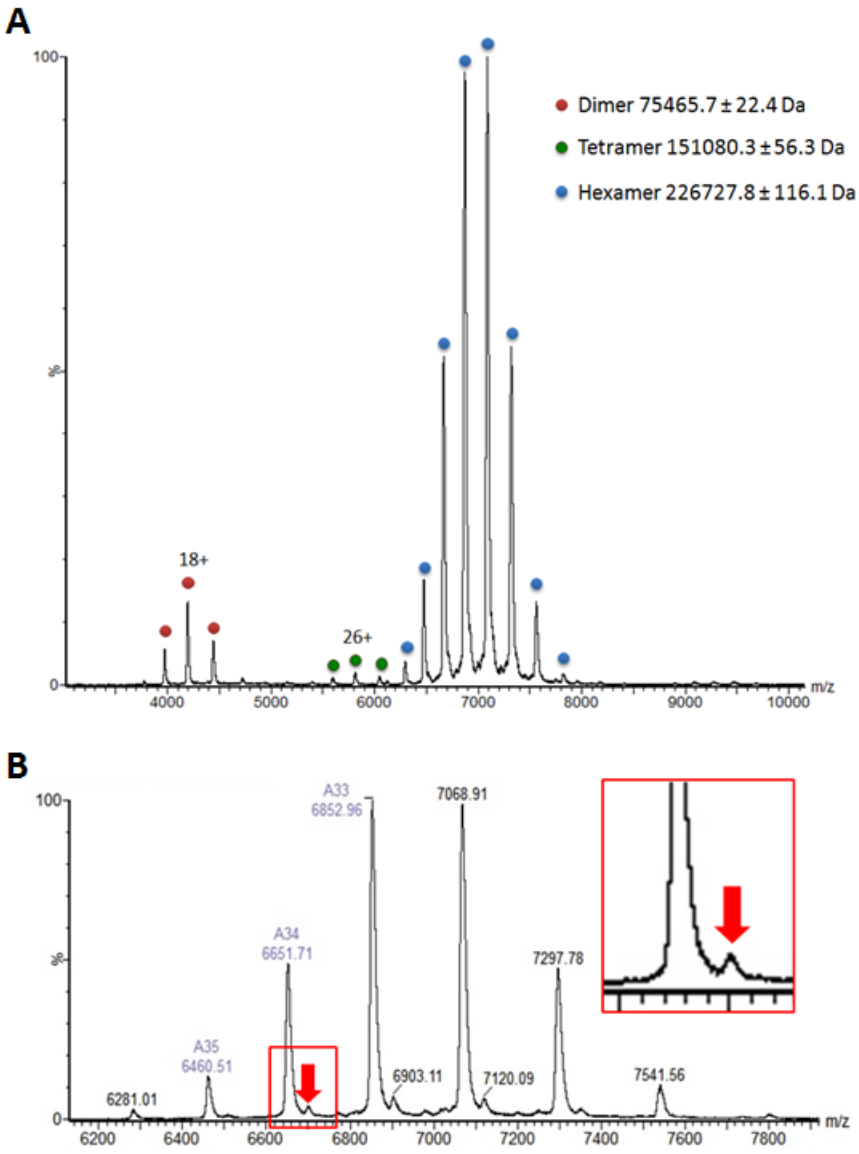


Figure 5.11: Native mass spectrum of ExsG. A - typical full-range spectrum showing different oligomeric forms of ExsG. B - partial spectrum acquired at higher collisional energy showing the resolved main and "shoulder" peaks, representing 226 and 228 kDa species, respectively. Red arrow indicates one of the "shoulder" peaks. Data acquired and analysed by Jun Yan.

assembly of the homohexamer. Given that all HKs characterised so far are homodimeric, an assembly involving a trimer of dimers was plausible; under such scenario the autocatalytic activity of ExsG would reside within a pair of subunits as in canonical HK cores.

To test if ExsG dimers are indeed the structural units, protein samples were prepared at several ammonium acetate concentrations, ranging from 300 mM to 800 mM. Ammonium acetate is the most commonly used buffer in ESI MS but higher concentrations can disrupt non-covalent electrostatic interactions between the constituent subunits of higher-order structures [157], which can be used to establish the assembly pathway for macromolecular complexes. The spectra acquired at each concentration indicated that the proportion of ExsG dimers increases with respect to the hexamers when increasing ammonium acetate concentration were used (Fig. 5.12). This suggests that at higher salt concentration the electrostratic interactions holding ExsG hexamer together were disrupted, and that the basic forming unit of the higher-order structure is a dimer; therefore ExsG is most likely a trimer of dimers.

5.3 Structural studies of ExsG

ExsG is unusual as it contains the relatively uncommon HWE HK core and exhibits both HK and RR activities. In addition, as described in the section above, contrary to the canonical HKs it was shown that ExsG forms a homohexamer with the assembly pathway involving dimeric units. Various approaches were undertaken to determine the three-dimensional structure of ExsG.

5.3.1 Crystallisation attempts

WT ExsG, ExsG-H151N and ExsG-HK were all subjected to crystallisation trials using sparse-matrix screens. Each protein was buffer exchanged into a low ionic strength (30 - 50 mM NaCl) Tris-Cl pH 8.2 buffer and prepared at several concentrations. Altogether about 700 crystallisation conditions were tested with protein. As no crystals were obtained, attempts were made to crystallise the proteins in the presence of magnesium chloride, ADP, AMP-PNP (an ATP analogue), or magnesium chloride with one of the nucleotides. Furthermore, full-length proteins were also subjected to crystallisation trials in the presence of beryllofluoride and magnesium chloride.

However, no crystals were observed, and it appeared that ExsG protein products frequently produced heavy precipitate, especially when present at higher concentrations.

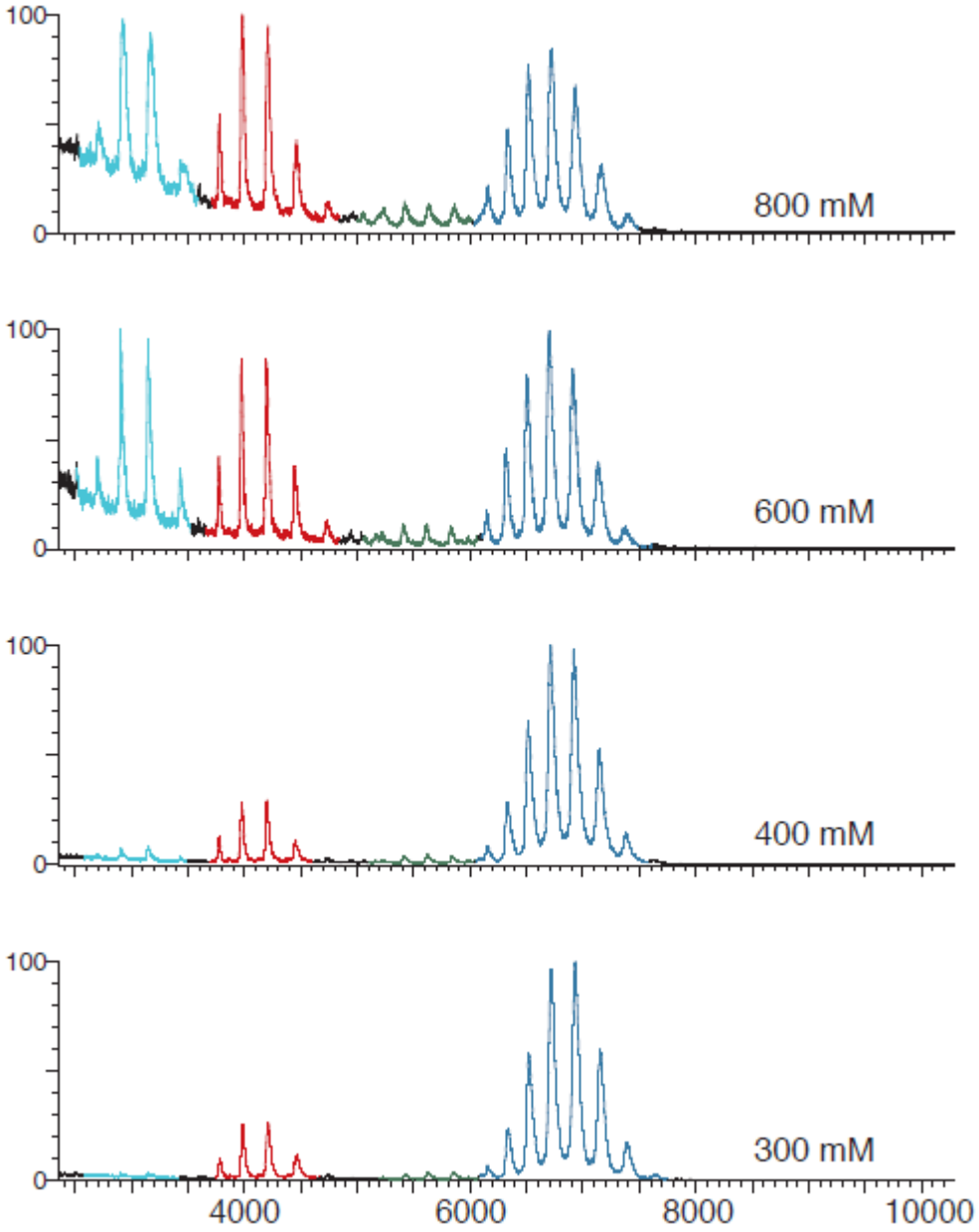


Figure 5.12: Determination of ExsG assembly pathway by ESI MS. The dark blue region of the spectrum represents the hexamer, green - tetramer, red - dimer, cyan - monomer. Data acquired and analysed by Jun Yan.

This could be caused by sample heterogeneity - although the covalently modified forms of ExsG were only detected in the case of the wild-type protein, the occurrence of some post-translational modifications in the other two proteins could not be excluded. ExsG protein products containing the HK core may also not be amenable to crystallisation due to the presence of highly flexible regions, such as the hinge region connecting DHp and CA domains in canonical HKs. Even though the catalytic domain of HWE HKs is significantly different from the canonical CA domains in terms of amino acid sequence, the overall mechanism of autophosphorylation, involving a mobile CA domain, could apply to HWE HKs. As explained in the previous section, it is likely that the structural unit of ExsG is a dimer, which suggests that the His-containing and catalytic domains of ExsG could be arranged and regulated in a similar fashion to classical proteins. This is explored in more detail in Section 5.4. The flexible linker conferring the mobility of CA domain seems to be the main reason for difficulties in crystallisation of entire HK cores, and it may also preclude the crystallisation of ExsG and its stand-alone HK core.

In the context of ExsG-HK, crystallisation failure may also be due to the fact that an 18-residue part of the interdomain linker, extending between the receiver domain and the HK core, was incorporated in that construct. As described in Section 5.4, these 18 residues and the first 4 residues of the HWE_HK domain are arranged in a heptad pattern and are thus likely to form a coiled-coil. However, in the absence of the receiver domain and the remaining part of the linker the stretch of residues N-terminal to the HK core may be disordered, thus precluding crystallisation.

5.3.2 Negative-stain electron microscopy (EM) studies

The three-dimensional structure of ExsG or its HWE HK core could not be obtained through X-ray crystallography. As the native state of the wild-type protein was shown to be a homohexamer with the molecular weight greater than 150 kDa, it seemed potentially suitable for EM studies. This technique allows determination of low-resolution structures of macromolecules that are larger than 150 kDa.

Protein samples were prepared at several dilutions, from 0.1 to 1 mg/ml. The staining procedure and materials used are described in Materials and methods, Section 2.11. Staining with uranyl acetate yielded high contrast but the protein appeared heterogeneous (Fig. 5.13), probably due to aggregation. Based on the observation that ExsG is prone to aggregation at pH lower than 8 (see Section 5.2.4.1), it seemed likely that the low pH of uranyl acetate (∼4 - 5) had a negative effect on the protein. In fact, the theoretical isolelectric point of ExsG (5.1, as predicted by the ProtParam online

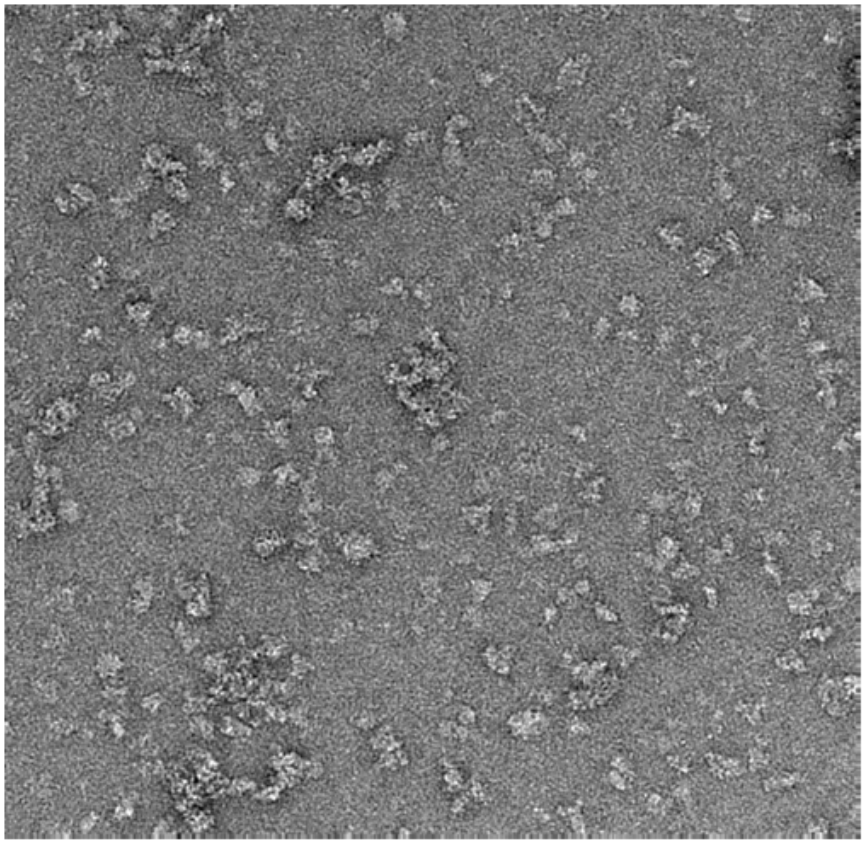


Figure 5.13: Electron micrograph of WT ExsG stained with uranyl acetate. Magnification 42000x, ExsG concentration ∼0.1 - 0.2 mg/ml.

tool[1]) is similar to the pH of uranyl acetate, which further suggests that the protein may have undergone aggregation as the electrostatic interactions were disrupted. Therefore two tungsten-based stains, NanoW^{TM} (pH ∼7) and phosphotungstic acid (PTA, pH adjusted to 7.6), were used to stain ExsG. As PTA is unstable at alkaline pH and does not possess the fixative properties of uranyl acetate, PTA solution and the stained carbon grids could not be stored and had to be prepared fresh for every EM experiment.

Even though the contrast achieved with tungsten stains is lower than with uranyl acetate, tungsten is characterised by a smaller grain which allows elucidation of finer features. Staining with NanoW^{TM} did not produce good enough contrast and, similarly to uranyl acetate, it seemed to induce aggregation. PTA pH ∼7.6 yielded better contrast and slightly reduced sample heterogeneity (Fig. 5.14); however, the protein still appeared relatively heterogenous in shape and size. A recurrent flower-like shape was observed, with the diameter estimated to be 10 - 15 nm, which could correspond to a 6-membered ring. Given that a canonical HK core (*Thermatoga maritima* HK853,

[1]http://web.expasy.org/protparam/

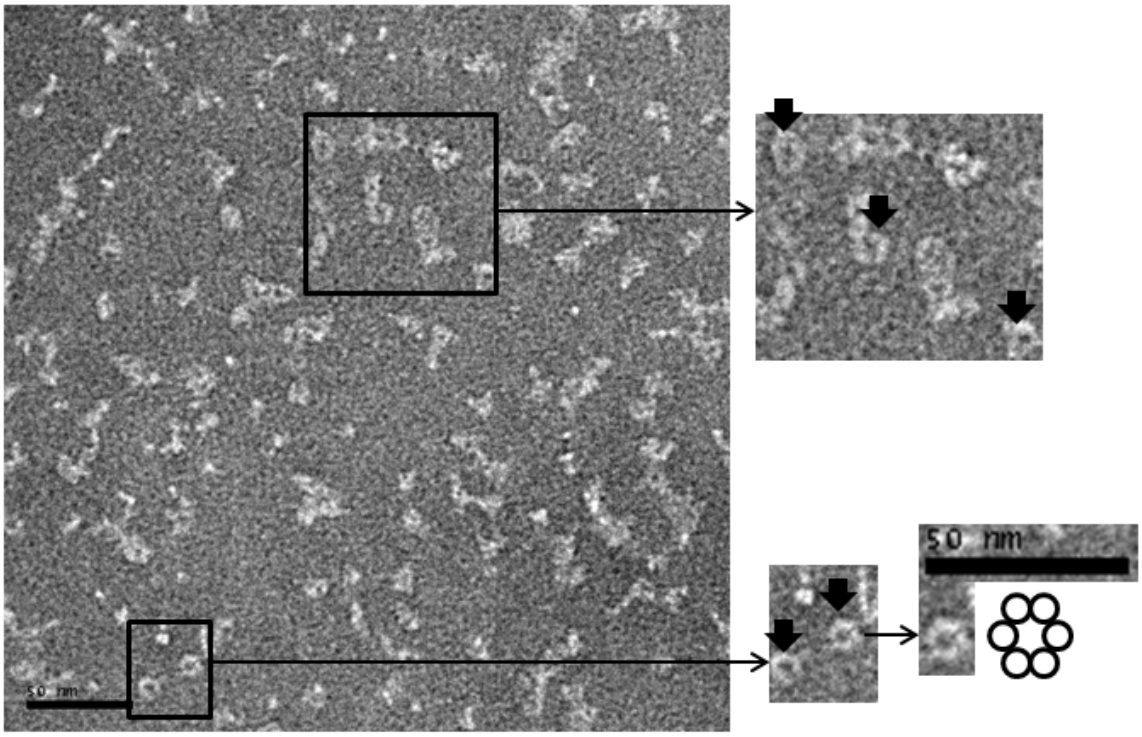


Figure 5.14: Electron micrograph of WT ExsG stained with PTA pH 7.6. Magnification 52000x, ExsG concentration 0.1 - 0.2 mg/ml. The flower-like structures are indicated with black arrows on magnified parts of the image. One of them is aligned with the scale bar and the hypothetical subunit arrangement (hexameric ring) is shown next to it.

PDB ID 3DGE) reaches about 8 - 8.5 nm in the longest dimension, and ExsG contains an additional domain (N-terminal receiver), this ∼10 - 15 nm flower-like structure could represent ExsG homohexamer. The constituent subunits may also be separated by a solvent channel, visible as a dark spot at the centre, which would add to the diameter.

Addition of 1 mM magnesium chloride and extensive phosphorylation of ExsG receiver domain with phosphoramidate prior to staining and visualisation did not seem to have a positive impact on the homogeneity of the sample. Therefore further experiments were not attempted.

5.4 Secondary structure prediction and comparison to canonical HKs

As the attempts to determine the three-dimensional structure of ExsG failed, computational approaches were undertaken to establish if low sequence identity implies lack of structural similarity between HWE and canonical HKs. Secondary structure prediction

for the full-length protein (Fig. 5.15) indicated that after the fifth helix, which constitutes the last element of the receiver domain, there is a long (∼53 amino acids) helix incorporating the active site histidine (His151), followed by another relatively long helix (23 residues). These two helical elements seemed equivalent to the two helices forming the DHp domain in canonical HKs, and could therefore also dimerise and form a helical bundle; as described in Section 5.2.4.2, despite the fact that ExsG exists as a hexamer, the basic structural unit is a dimer. Interestingly, the presence of the long helical linker connecting the HK core to the receiver domain seemed analogous to the signalling helix, a frequent structural motif usually connecting sensory regions with signalling modules [28].

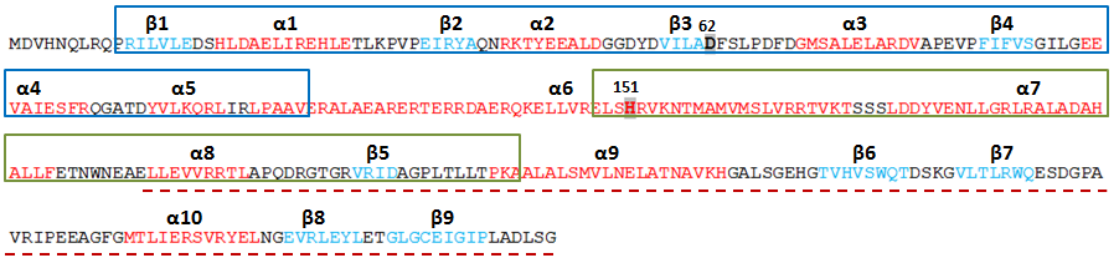


Figure 5.15: Secondary structure prediction by PSIPRED for full-length ExsG. β-strands are indicated by blue font, helices - by red font, the phosphorylatable residues (Asp62 and His151) are highlighted in grey. The regions recognised as domains by SMART are framed in blue (receiver) and green (HWE_HK). Red dashed line delineates the putative ABD (see text for explanation).

25 ExsG-like proteins, with N-terminal receiver domains and C-terminal HK cores, were found to contain the signalling helix between the receiver domain and the HK core [28]. This structural motif exhibits heptad periodicity which suggests that two signalling helices dimerise to form a coiled-coil. Within the identified heptad pattern the amino acid residues are labelled as 'abcdefg'. The 'a' and 'd' positions of the heptad pattern are typically occupied by hydrophobic residues which mediate coiled-coil interactions. In many HKs the signalling helix directly precedes the DHp domain and merges with helix α1, which contains the phosphorylatable histidine residue. To identify if the long helical linker within ExsG sequence could constitute the signalling helix, ExsG sequence was analysed using the programme COILS [117] that tests for the presence of heptad repeats. It appears that 22 residues, including the active site histidine (His151), were arranged in a heptad pattern (Fig. 5.16). Typical signalling helix involves five heptad repeats and contains several conserved amino acid motifs [28]; ExsG linker exhibited three repeats and contained only one discernible common sequence motif, ERT (Glu, Arg, Thr). However, even the ERT motif might be coincidental as the arginine was

assigned position 'e', whereas in a heptad repeat of the signalling helix it occupies position 'd'. Furthermore, half of the assigned 'a' and 'd' positions were occupied by charged residues, suggesting that, in contrast to typical coiled-coils, interactions between the two helices may involve salt bridges.

Taking all the above into consideration, the linker region between the receiver domain and HK core of ExsG does not seem to constitute the signalling helix, but is likely to form a coiled-coil based on the heptad periodicity, and may lead directly into helix α1 of the HWE_HK domain. Potential mechanisms of interdomain signal transduction that involve the helical linker are described in the Conclusions and discussion. In analogy to ExsG, the signalling helix was not identified within the interdomain linker in the other NT-26 hybrid RR (HK/RR2677), which exhibits identical domain composition and arrangement. However, it was found within the linker connecting the N-terminal receiver domain with the HisKA core in a *Bradyrhizobium japonicum* HK (HK3656, UniProt entry Q89QY6) and the other NT-26 hybrid RR, based on the presence of at least five heptad repeats and all the signature residues. Therefore signal transmission in hybrid RRs containing a canonical core seems to involve the signalling helix, while intramolecular signalling between the receiver domain and a HWE core relies on a potentially distinct mechanism.

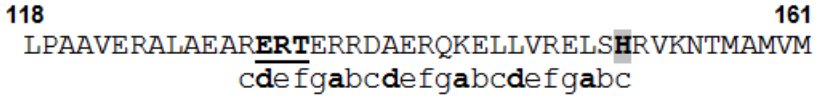


Figure 5.16: Region of ExsG interdomain linker predicted to form a coiled-coil by the COILS programme. The a-g positions within heptad repeats which were predicted with a probability > 0.9 are shown below the sequence. Positions 'a' and 'd', forming the interactions between helices, are in bold. His151 is highlighted in grey and the putative signature motif of signalling helix (ERT) is underlined.

As mentioned above, the two helices within the HWE_HK domain seem to be equivalent to the two helices forming the DHp domain of canonical HKs. According to the secondary structure prediction (Fig. 5.15), there are three α-helices and five β-strands C-terminal to the DHp-like region of ExsG. These elements are likely to form the ATP-binding domain (ABD), responsible for binding and hydrolysis of the nucleotide. Curiously, despite the fact that this C-terminal portion of ExsG shares little sequence similarity with HATPase_c domains (see Chapter 4), it exhibits a similar order of secondary structure elements as HATPase_c. Both the high-resolution structure (PDB ID 3EHG [158]) and the secondary structure prediction of the CA domain of a thermosensor HK, DesK, indicate the same topology. The ABD of ExsG is also

topologically similar to the CA domain of HK853 (PDB ID 3DGE [32]).

To further test the possibility that the entire HK core of ExsG is structurally analogous to that of canonical HKs, the sequence of ExsG-HK was scanned against the CATH database [159]. The entire C-terminal portion, comprising three helices and five β-strands, was recognised as a two-layer α/β sandwich fold and classified as a member of CATH superfamily 3.30.565.10 (sensor protein-like domain). This superfamily encompasses HKs, topoisomerases and the DNA mismatch repair protein MutL, all of which belong to the GHKL superfamily of ATPases. Therefore, despite the fact that ExsG ABD is recognised by SMART and Pfam as a HATPase_c domain with only a borderline significance value (8.99), it most likely adopts the same two-layer sandwich fold characteristic of GHKL ATPases [59]. Furthermore, pGenTHREADER fold recognition tool [120] identified several HK cores as the closest structural homologues of ExsG HK core, with HK853 (PDB ID 2C2A [62]) as the highest scoring homologue. Notably, ExsG shares 18.4% sequence identity with HK853 and 16.8% with NT-26 BphP1 in terms of the HK core.

In conclusion, it appears that the structure of the HWE HK core of ExsG is analogous to the HisKA-type core, and may form a similar four-helix bundle upon dimerisation. The main structural differences are most likely present within the ATP-binding domain; HWE HK cores exhibit a potentially smaller ATP lid and lack the F-box (see Chapter 4 and [60]). It remains unknown how the hexameric structure is assembled but according the results presented above (Section 5.2.4.1) oligomerisation is mediated via the HWE HK core. Importantly, given that the ExsG ABD is likely to constitute the entire C-terminal portion starting with helix α8 (Fig. 5.15), the HWE_HK domain boundaries, as defined in SMART database (ID SM00911), may be incorrect. According to SMART domain annotation, the HWE_HK domain extends into the putative catalytic region, encompassing two of its elements (α8 and β5).

5.5 Summary

With the relatively uncommon HWE HK core and dual HK/RR activity, ExsG was an interesting target for structural characterisation. Even though structure determination attempts proved unsuccessful, using several biophysical techniques it was herein demonstrated that ExsG exists primarily as an ellipsoid homohexamer, which is assembled from protein dimers. Furthermore, the formation of the higher-order structure was shown to be promoted by the HWE HK core of ExsG, while the stand-alone receiver domain is insufficient for hexamerisation. The possibility that the interdomain helical

linker extending into the the first α-helix of ExsG HK core constitutes the signalling helix was rejected, but it was shown that a portion of that region has a propensity to form a coiled-coil structure.

Notably, the arrangement of the predicted secondary structures within the HWE core of ExsG was demonstrated to be analogous to canonical HK cores, and the ATP-binding domain of ExsG seems to belong to the GHKL ATPase superfamily. It was also shown herein that the closest structural homologues of ExsG HK core are canonical HKs. Together with the fact that the structural unit of ExsG is a dimer, these results suggest that HWE HK cores may be structurally, and mechanistically, related to canonical HKs.

Chapter 6

Determination of the signalling role of ExsG

As described in Chapter 4, ExsG contains two phosphorylatable residues, Asp62 and His151, and exhibits both HK and RR activities. This raised the question concerning regulation - the activity of the HK core could be regulated by phosphorylation of the receiver domain, and *vice versa*. It seemed likely that, in analogy to RRs containing C-terminal output domains, phosphorylation of the N-terminal receiver domain would affect the autokinase activity.

The activity of the isolated receiver domain of ExsG (ExsG-REC) was demonstrated and is described in Chapter 4. The amount of the radioactive signal transferred from BphP1-HK onto full-length ExsG and ExsG-REC was comparable (see Fig. 4.19), suggesting that the presence of the C-terminal HK core is not necessary for RR activity, i.e. for phosphorylation of the receiver domain. Similarly, as described in Chapter 5 (Section 5.2.1), the stand-alone HK core of ExsG (ExsG-HK) retained autokinase activity, although the level of activity was about 5-fold lower than that of WT ExsG (see Fig. 5.3). The latter observation could be explained by the reduced stability of the truncated protein - the ExsG-HK construct incorporated a part of the interdomain linker, which in the absence of the remaining N-terminal portion of the protein may not fold correctly. Alternatively, removal of the receiver domain might have stabilised the inactive form of the HK core, which would imply that the HK core is under the control of the receiver domain. Several experimental approaches were applied herein to establish which domain controls the activity of the other one in a phosphorylation-dependent manner, and to verify if ExsG can exist as a doubly phosphorylated protein (with both aspartate and histidine phosphorylated).

ExsG seemed likely to act as a RR, with autokinase activity as the signalling output,

and was shown to exist as a homohexamer in contrast to canonical HKs (see Chapter 5). It was therefore possible that it might exhibit novel mechanisms of signal transmission between the receiver domain and the HK core. Native mass spectrometry and ion mobility were employed here to study the conformational changes associated with nucleotide binding and phosphorylation of the receiver domain.

6.1 Determination of the effect of receiver domain phosphorylation on HK core activity

6.1.1 Beryllofluoride test

The initial approach aimed to establish if ExsG autokinase activity is controlled by the receiver domain in a phosphorylation-dependent manner involved beryllofluoride, which was shown to mimic the effect of phosphorylation of receiver domains [160]. The compound is formed by mixing beryllium chloride with an excess of sodium fluoride (6-fold excess was used herein). WT ExsG was incubated for one hour in the "hot" reaction buffer containing varying concentrations of beryllofluoride, after which the samples were passed through the chromatography spin columns to remove the excess of radioactively labelled ATP. The resulting autoradiograph and the plot of the normalised PSL signal against beryllium chloride concentration indicated that ExsG autokinase activity decreased with increasing concentration of the compound (Fig. 6.3).

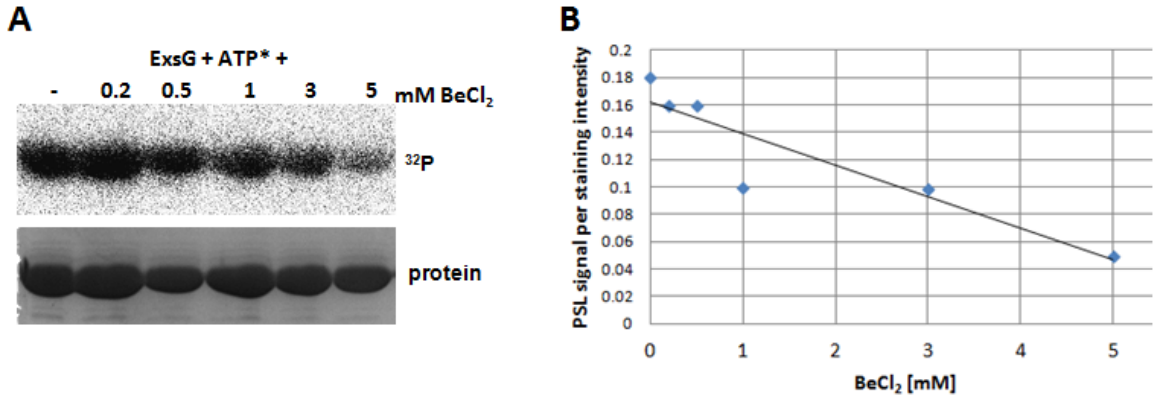


Figure 6.1: The effect of different beryllofluoride concentrations on ExsG autokinase activity. A - autoradiograph and the corresponding SDS-PAGE gel; B - plot of the PSL signal normalised per Coomassie blue staining intensity versus $BeCl_2$ concentration.

To confirm that beryllofluoride did not directly interfere with autophosphorylation, for example by binding to the ATP-binding domain and thus preventing ATP hydrolysis, both WT ExsG and ExsG-HK were subjected to the autokinase assay in the

presence of different concentrations of the compound. As evident from Figure 6.2, the extent of phosphorylation of WT ExsG in the reaction containing a high beryllofluoride concentration was significantly reduced, while the autokinase activity of ExsG-HK seemed to be unaffected. It therefore appeared unlikely that beryllofluoride inhibited ExsG autophosphorylation by interacting directly with the HK core.

It was possible that beryllofluoride decreases the apparent autokinase activity of ExsG in yet another way - it could diminish the number of phosphorylated molecules by reducing the stability of the phosphohistidine product. To test that, WT ExsG was allowed to autophosphorylate, purified from the unreacted ATP and incubated for two hours with or without beryllofluoride. Even though the autokinase activity of ExsG-HK did not seem to be affected by the compound, the stability of ExsG-HK phosphohistidine was also examined in the same fashion. Beryllofluoride had no detectable impact on the stability of ExsG phosphohistidine, both in the context of the full-length protein and the stand-alone HK core (Fig. 6.3). Taking all the above results into consideration, beryllofluoride acted specifically on full-length ExsG, and therefore the evident decrease in autokinase activity most likely resulted from the binding of beryllofluoride to the receiver domain. As the compound did not affect phosphohistidine stability, it transpired that phosphorylation of the receiver domain inactivates the HK core.

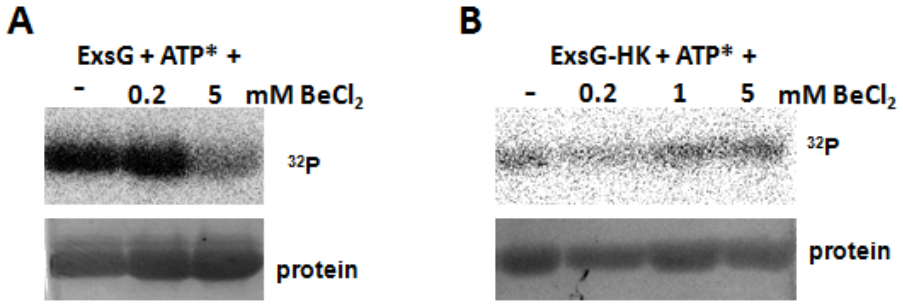


Figure 6.2: Comparison of the effect of beryllofluoride on autokinase activity of WT ExsG and ExsG-HK. Autoradiographs show the results of WT ExsG (A) and ExsG-HK (B) autokinase assays in the presence of varying beryllofluoride concentrations. Stained gels indicate protein levels.

However, the autoradiographs and the corresponding SDS-PAGE gels indicated the presence of an additional protein species associated with a significant amount of radioactive signal (indicated by arrows in Fig. 6.3). It was more pronounced above WT ExsG protein bands, which could be due to a higher concentration of WT ExsG used in the experiment. The species was likely to represent protein aggregates as it was retained on the border of the stacking and resolving gels, and it was not observed in the

previous beryllofluoride experiments because there the samples were passed through the chromatography spin columns prior to loading onto the gel. Here the reactions were stopped by adding the loading dye directly.

Given that these high molecular weight species were linked to the presence of beryllofluoride, it was likely that the compound may affect the native state of ExsG, possibly by interacting with some of the components of the buffer. Therefore several test incubations were set up where a number of parameters were examined for their contribution to beryllofluoride-induced aggregation. Non-radioactive SDS-PAGE analysis suggested that beryllofluoride promoted the formation of aggregates independently of the presence of magnesium chloride and the reducing agent, while beryllium chloride and sodium fluoride on their own did not induce aggregation (Fig. 6.4). High molecular weight species was also observed in the case of BphP1-HK that was mixed with beryllofluoride prior to loading on the gel (Fig. 6.4C). Similarly, aggregation occurred primarily in samples containing only beryllofluoride; when ATP was also present in the reaction, the high molecular weight species was not observed. Therefore, it is likely that the compound may interact with ATP, which reduces its effective concentration, thus preventing it from inducing protein aggregation. However, it remains unknown why such high amount of radioactive signal was associated with the aggregated protein in the radioactive assays (see Fig. 6.3).

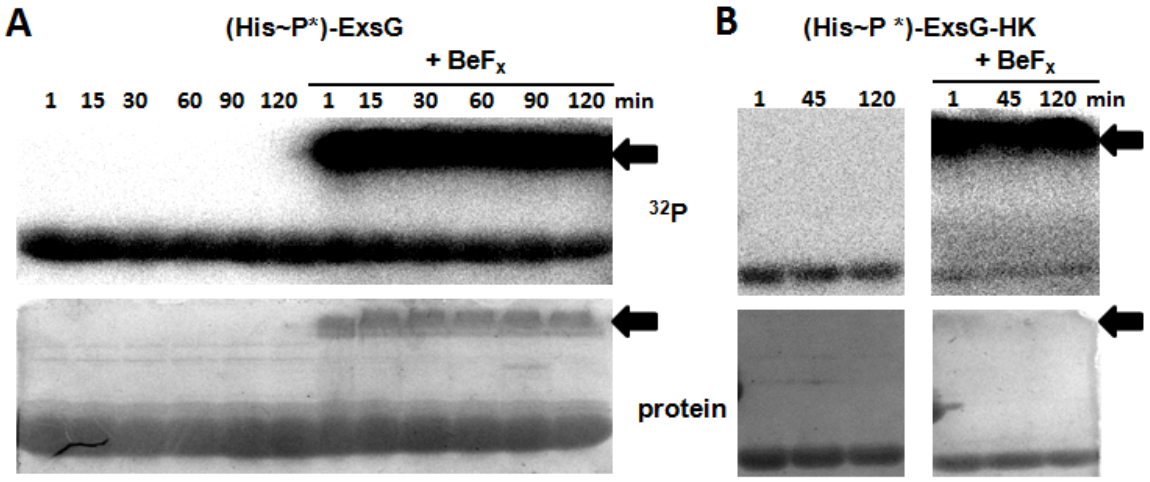


Figure 6.3: The effect of beryllofluoride on the stability of WT ExsG and ExsG-HK phosphohistidines. Phosphorylated WT ExsG (A) and ExsG-HK (B) were cleaned from ATP and incubated with or without beryllofluoride (5 mM $BeCl_2$, 30 mM NaF). Arrows indicate the high molecular weight aggregates.

Notably, removal of the reducing agent both from the reaction and the loading dye resulted in the appearance of ExsG dimers (75 kDa, indicated by an arrow in Fig.

6.4). They probably form via disulfide bridges between denatured monomers - ExsG harbours one cysteine residue within its sequence.

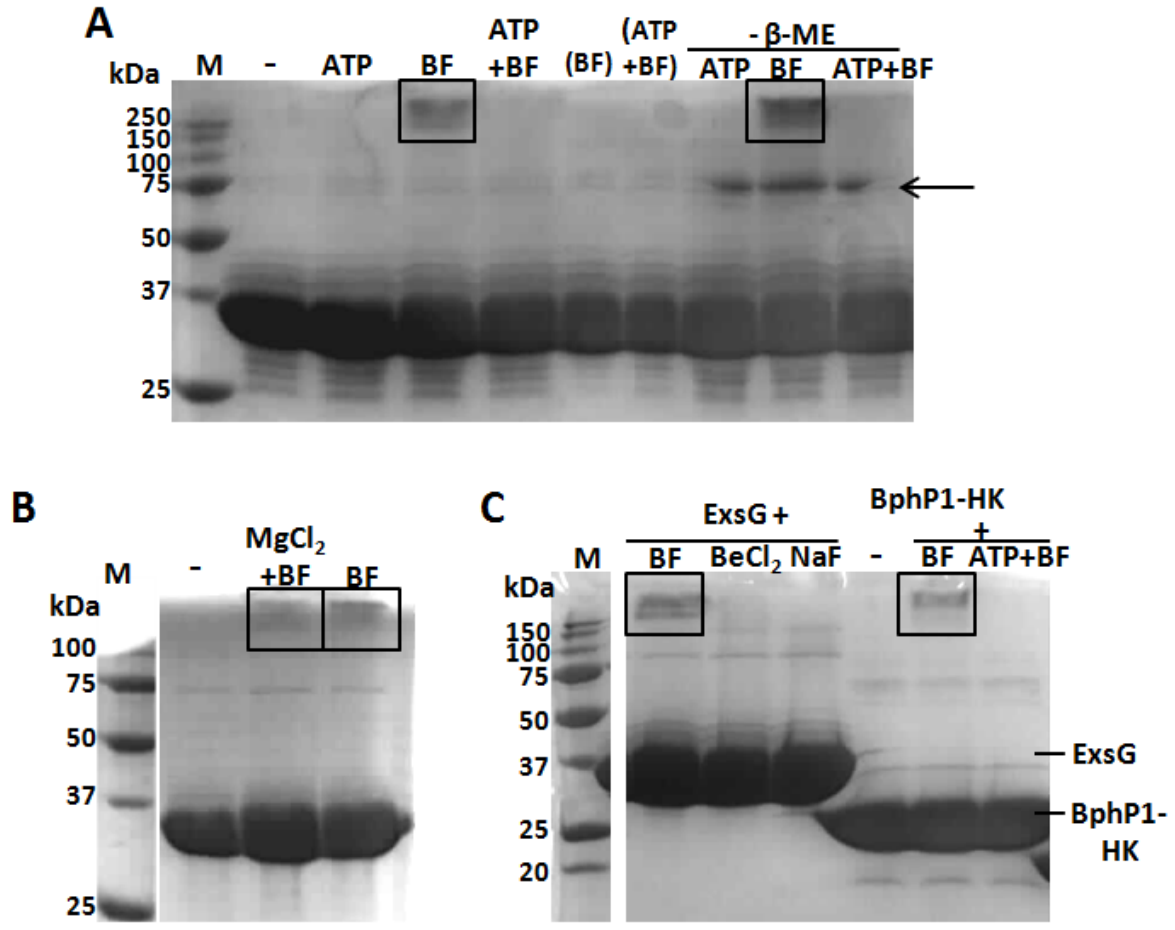


Figure 6.4: Determination of the parameters contributing to beryllofluoride-mediated protein aggregation. SDS-PAGE analyses of WT ExsG (A, B, C) and BphP1-HK (C) incubated with, without, or with a combination of, ATP and beryllofluoride (BF). Three ExsG reactions (A) were performed in the absence of the reducing agent (β-ME), which was also absent from the loading dye for this set of reactions. One reaction was performed with BF and 10 mM magnesium chloride (B). Brackets indicate reactions purified with the chromatography spin columns prior to electrophoresis; black arrow indicates ExsG dimers; aggregate bands are framed. Concentrations of ATP, $MgCl_2$ and β-ME were the same as in standard radioactive assays, BF constituted of 5 mM $BeCl_2$ and 30 mM NaF.

6.1.2 Phosphoramidate (PA) test

The results described above implied that phosphorylation of the receiver domain down-regulates autokinase activity. However, as a high concentration of the compound was associated with protein aggregation, the conclusions based on the beryllofluoride expe-

riments were verified using protein with covalently attached phosphate. PA was used here instead of the cognate HK of ExsG, BphP1. Following the phosphorylation of the receiver domain, the compound could be conveniently removed from the reaction using the chromatography spin columns, which eliminated the possibility of its potential interactions with the protein, ATP and other components of the "hot" reaction buffer. An ExsG variant lacking the phosphorylatable aspartate, ExsG-D62N, was used as a control in some of the experiments described below.

6.1.2.1 Radioactive assay using standard SDS-PAGE

In the first experiment, WT ExsG and ExsG-D62N were incubated with PA for approximately two hours to allow efficient aspartate phosphorylation, after which the compound was removed. Autokinase reactions of ExsG variants pre-incubated with PA, as well as proteins not exposed to PA, were initiated by addition of the "hot" reaction buffer, and the extent of phosphorylation was monitored over the course of 1 hour. The results demonstrated that ExsG-D62N, irrespective of whether it was pre-incubated with PA or not, exhibited similar levels of autokinase activity (Fig. 6.5). In contrast, the extent of phosphorylation of WT ExsG which was previously incubated with PA was significantly lower than that of untreated protein - even after 1 hour the radioactive signal was hardly detectable. This observation was in line with the results from the beryllofluoride experiment; phosphorylation of the receiver domain seemed to inactivate the HK core.

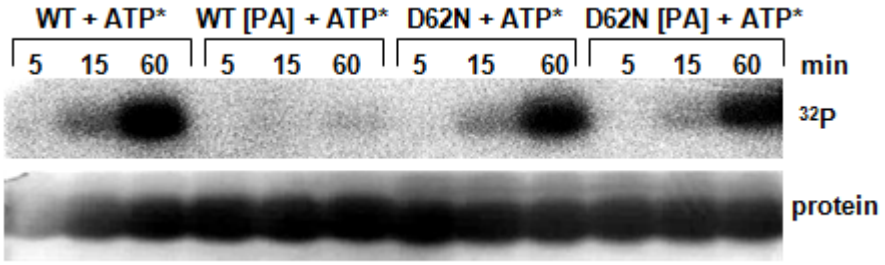


Figure 6.5: The effect of phosphorylation of ExsG receiver domain on autokinase activity determined using standard SDS-PAGE. Autoradiograph and corresponding SDS-PAGE gel show the results of WT ExsG and ExsG-D62N autokinase assays. [PA] indicates proteins which were pre-incubated with PA for 2 hours.

Beryllofluoride experiments indicated that the compound did not have any detectable influence on the stability of pre-formed phosphohistidine (Section 6.1.1). In order to verify that phosphorylation of the receiver domain did not affect the half-life of the phosphohistidine, ExsG was allowed to autophosphorylate in the "hot" reaction buffer

and, after removing unincorporated nucleotide, PA was added. Samples were taken over the course of one hour and analysed by SDS-PAGE and autoradiography. As evident from Figure 6.6, there was no observable decrease in the level of histidine phosphorylation, which confirmed that aspartate phosphorylation of the receiver domain prevents kinase core autophosphorylation, but does not reduce the stability of existing phosphohistidines.

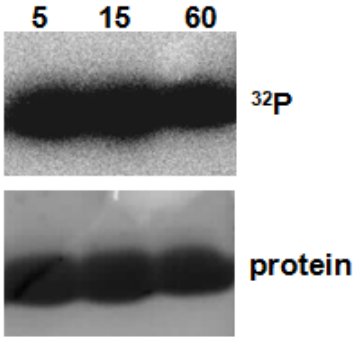


Figure 6.6: Effect of phosphorylation of the receiver domain of ExsG on phosphohistidine stability

6.1.2.2 Radioactive phosphoprotein affinity gel electrophoresis

To identify if (Asp∼P)-ExsG molecules can be phosphorylated on the histidine, radioactive autokinase assay was coupled with phosphoprotein affinity gel electrophoresis. Aspartate phosphorylation was previously shown to produce an upshifted ExsG band, and it was expected that His151 phosphorylation would result in a similar mobility shift, while molecules phosphorylated on both of these residues would appear as a third, further upshifted band on the gel. Preliminary non-radioactive tests, however, indicated that the migration of (His∼P)-ExsG through the Phos-tag^{TM}/Mn^{2+} gel is not affected, in contrast to (Asp∼P)-ExsG and BphP1-HK∼P (Fig. 6.7). Although, as previously demonstrated, aspartate phosphorylation resulted in as an upshifted ExsG protein band, histidine phosphorylation had no effect on ExsG gel migration. By means of the radioactive autokinase assay it was confirmed that the radioactive signal was confined to the only ExsG band visible on the gel (Fig. 6.8). The same behaviour was observed in the case of ExsG-HK (Fig. 6.8); apparent lack of separation of (His∼P)-ExsG suggested that ExsG phosphohistidine is not bound by the Phos-tag^{TM}/Mn^{2+} moiety.

According to Phos-tag^{TM}acrylamide manufacturers, most phosphoproteins are separated using gels containing the Phos-tag^{TM}/Mn^{2+} moiety; in fact, one of the examples of

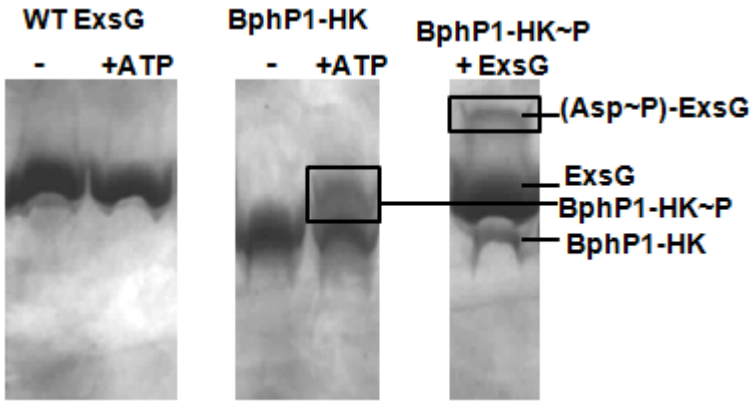


Figure 6.7: Determination of phospho-ExsG and phospho-BphP1-HK mobility shifts by non-radioactive phosphoprotein affinity gel electrophoresis. Phosphoproteins are framed.

successful phosphoprotein separation, shown in the protocol provided with Phos-tag^{TM} acrylamide, involves an unnamed histidine kinase. However, cases where resolution does not occur have been reported [161]. The replacement of manganese with zinc, which seems to provide better separation, has not been tested herein. Curiously, as a mobility shift occurred in the case of phosphorylated BphP1-HK (Fig. 6.8), the inability of the Phos-tag^{TM}/Mn^{2+} moiety to bind (His∼P)-protein may be specific to HWE HKs or ExsG alone.

For clarity, in the context of ExsG the "lower" (main) band refers to unphosphorylated and/or (His∼P)-ExsG, while the "upper" (upshifted) band represents (Asp∼P)-ExsG. Here an attempt was made to determine if the upper band can contain (His∼P)-ExsG as well as protein phosphorylated on the aspartate - if double phosphorylation was to occur, the radioactive signal resulting from ExsG autophosphorylation (His∼P*) should be observed within the upshifted band (Asp∼P).

In the first attempt two autokinase assays were set up, one containing untreated ExsG and the other containing protein that was pre-incubated with PA for approximately 30 minutes and subsequently purified using chromatography spin columns. Samples taken out of each reaction over the course of 1 hour were cleaned from excess "hot" ATP by passing them through the chromatography spin columns again, prior to addition of the loading dye; this was aimed to reduce the background radioactive signal on the autoradiograph. However, the elutions obtained varied in volume thus affecting protein concentration. The resulting autoradiograph and stained gel indicated that some of the samples loaded on the gel contained significantly higher concentration of the protein (Fig. 6.9). Furthermore, these highly concentrated samples exhibited an additional protein band positioned at the same level as (Asp∼P)-ExsG and associated with low levels of radioactivity; this band could represent one of the commonly observed cova-

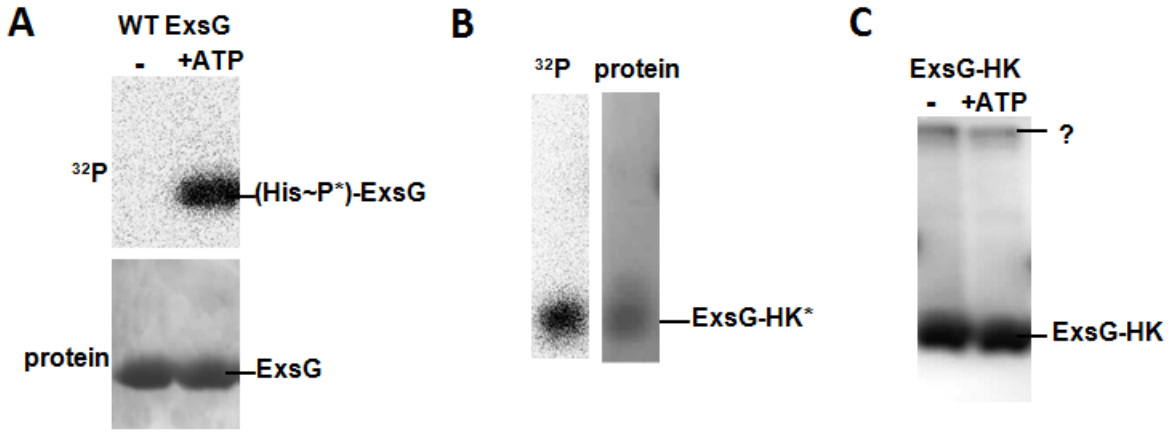


Figure 6.8: Verification of the position of (His∼P)-ExsG on Phos-tag^{TM}/Mn^{2+} SDS-PAGE gels. A - WT ExsG incubations with or without the "hot" reaction buffer. B - autokinase reaction of ExsG-HK showing the radioactive signal being confined to a single protein band. C - ExsG-HK incubations with or without the ATP buffer (non-radioactive). Question mark denotes a possible contaminant associated with this batch of ExsG-HK.

lently modified forms of ExsG (see Chapter 5), for example protein phosphorylated on Ser163. Nevertheless, a trend was evident - majority of the radioactive signal per lane was confined to the lower band. Despite the smearing of the protein bands, it was apparent that at least a quarter of ExsG molecules were phosphorylated on the aspartate after the incubation with PA, but almost all the radioactivity was detected within the lower band after the autokinase reaction. The extent of histidine phosphorylation was considerably lower in reactions containing a proportion of (Asp∼P)-ExsG than in those containing previously PA-untreated protein, and it seemed to correlate only with the protein content within the lower band.

To obtain further evidence that WT ExsG (Asp∼P)-ExsG is unlikely to be phosphorylated on the histidine, the experiment was repeated using ExsG populations with different proportions of molecules phosphorylated on Asp62. The protein used to set up the autokinase reactions was either untreated or pre-incubated with PA for various times (10, 45 or 120 min) after which the excess of PA was removed using chromatography spin columns. As before, the elution volumes varied due to the nature of the spin columns, resulting in different amounts of protein loaded on the gels. Therefore the radioactive signal had to be normalised to the protein content. Following the incubation with PA, autophosphorylation progress was monitored over 2 time points (15 and 60 min); this time samples taken out of the reactions were not further purified prior to loading onto the gel to avoid the aforementioned problem of varying protein concentrations. The results seemed to reflect the same trend (Fig. 6.10A); most of the radioactive signal was confined to the lower band. As the samples taken at 15 min

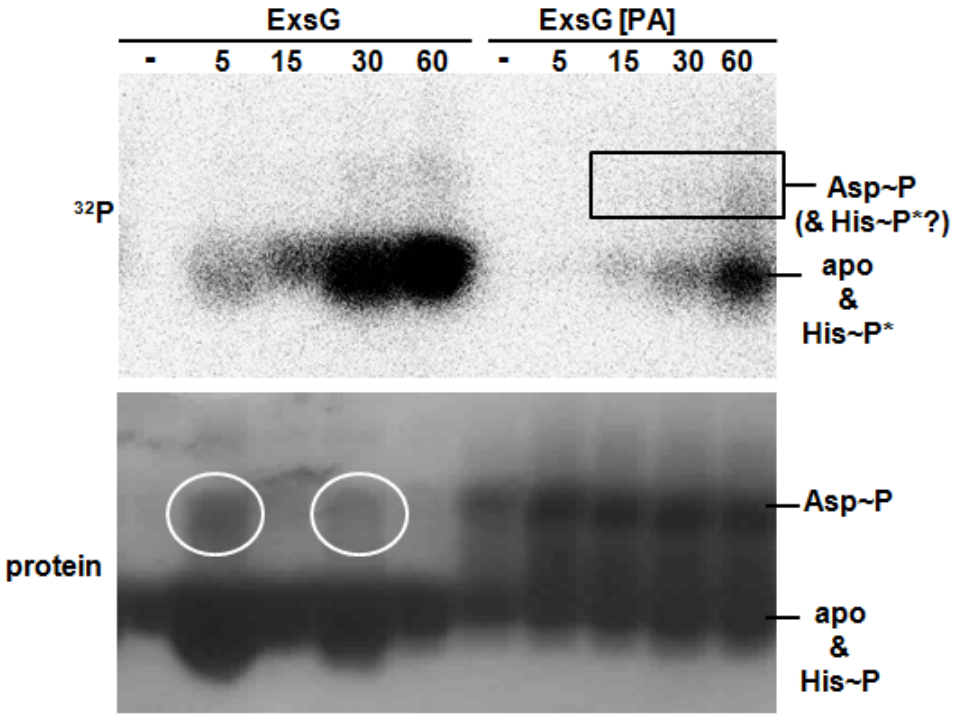


Figure 6.9: Preliminary test for (Asp∼P)-ExsG autophosphorylation. [PA] denotes protein pre-incubated with PA for 30 min after which the compound was removed. Apo - unphosphorylated protein. (Asp∼P)-ExsG bands exhibiting low levels of radioactivity are framed. Circles indicate the position of additional protein bands, overlapping with the (Asp∼P)-ExsG band, which most likely correspond to a covalently modified ExsG form (other than phosphorylated on His151 or Asp62).

exhibited background signal levels, only 1 hour autokinase reactions were considered herein and used to estimate the amount of signal within each band.

Low levels of radioactivity were observed within the upper band, as in the previous experiment (see Fig. 6.9), suggesting that some doubly phosphorylated protein molecules were present. However, it was evident that, even in the reaction where ∼57% of the molecules were phosphorylated upon aspartate (after 2 hours of incubation with PA), most of the radioactivity was associated with the lower band (Fig. 6.10B). Taking all the above results into consideration, molecules already phosphorylated upon Asp62 are significantly less likely to become phosphorylated upon His151, suggesting that phosphorylation of the receiver domain of ExsG abolishes autokinase activity.

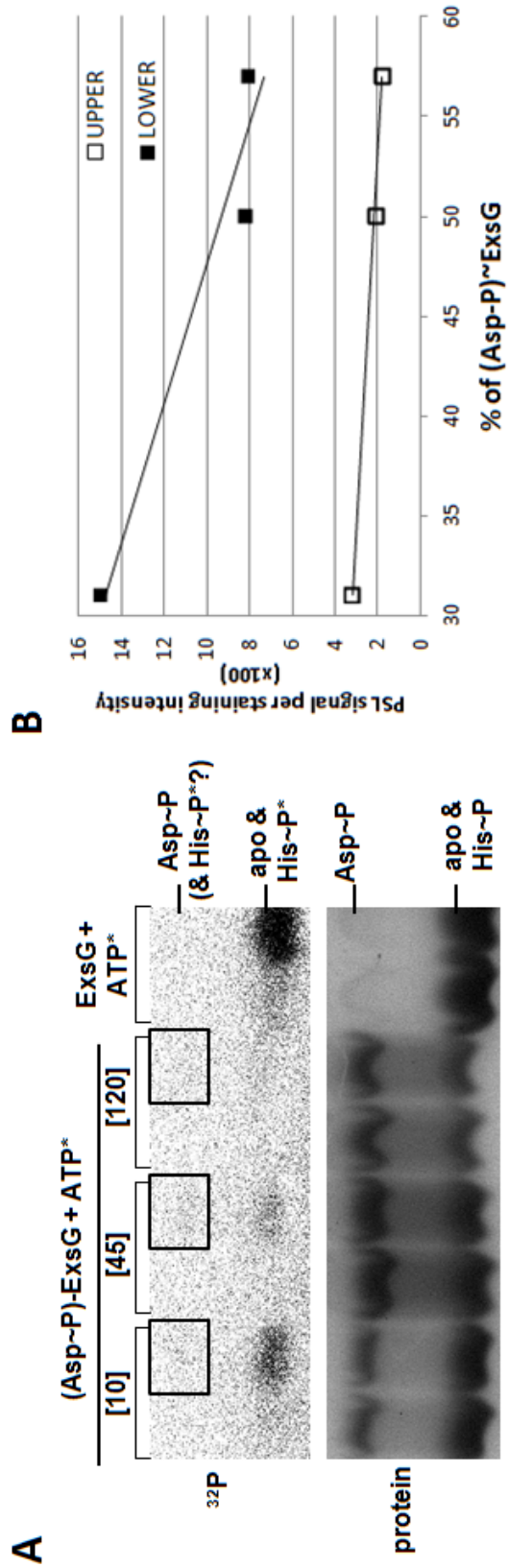


Figure 6.10: Determination of the possibility of (Asp~P)-ExsG autophosphorylation activity. A - autoradiograph and Phos-tag^{TM}/Mn^{2+} SDS-PAGE gel showing the results of autokinase assays of ExsG and (Asp~P)-ExsG. Numbers in square brackets indicate the length of pre-incubation with PA in minutes. Upper bands which exhibit detectable radioactivity levels are framed. Apo - unphosphorylated protein. B - plot of normalised PSL signal versus the fraction of ExsG molecules phosphorylated on Asp62. "Upper" and "lower" refer to the protein bands in (A); only the lanes containing 60 min reactions were used.

6.2 Determination of the effect of HK core phosphorylation on receiver domain activity

As shown in Chapter 4, the stand-alone receiver domain of ExsG retained a level of RR activity comparable to that of the full-length protein, indicating that the presence of the HK core is neither necessary for nor precluding the phosphorylation of the receiver domain. The results presented in the previous section suggested that some ExsG molecules can be phosphorylated upon both active site residues, but after phosphorylation of the aspartate the formation of the phosphohistidine is significantly less likely. This implied that the activity of the HK core is downregulated when the receiver domain is phosphorylated. However, a reverse scenario was also conceivable, whereby the RR activity of the receiver domain could be regulated by the HK core in a phosphorylation-dependent manner. This hypothesis was tested using radioactive phosphoprotein affinity gel electrophoresis; WT ExsG was first allowed to autophosphorylate in the "hot" reaction buffer and, after removal of ATP, it was incubated with PA for 1.5 hours.

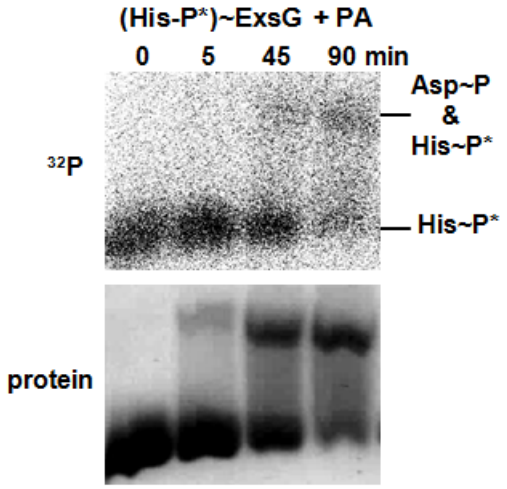


Figure 6.11: Test for the RR activity of (His∼P)-ExsG.

Figure 6.11 shows that both the protein and the radioactive signal were gradually transferred from the lower to the upper band, indicating that ExsG phosphorylated on the histidine could still be phosphorylated upon the aspartate. Therefore phosphorylation of the HK core does not abolish the RR activity of ExsG. Notably, however, the result of this experiment does not exclude the possibility that the RR activity of (His∼P)-ExsG may be reduced.

6.3 Ion mobility spectrometry - mass spectrometry (IMS-MS) analysis

Ion mobility spectrometry allows the separation of ionised molecules based on their mobility in a carrier gas, which depends on their collision cross-sections. The drift time of a molecule across the ion mobility chamber is therefore correlated with its size and compactness. Coupling IMS with ESI MS enables studying proteins and protein complexes in their native states, and conformational changes occurring upon ligand binding or covalent modifications can be determined by comparing the drift times. The data acquired using unbound (apo) ExsG (Fig. 6.12) indicated the presence of two distinct species which were observed in the case of protein ions containing 30 or more charges. These species were likely to correspond to distinct conformations of ExsG, exhibiting different stabilities under increased stress conditions (high charge densities).

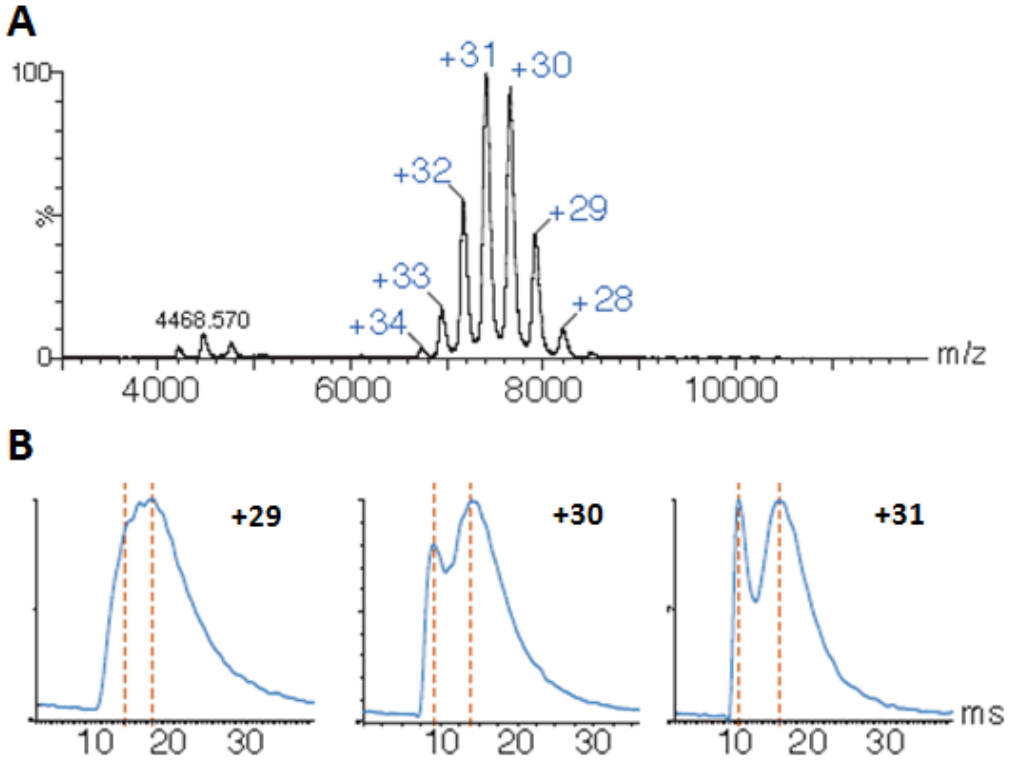


Figure 6.12: Analysis of apo-ExsG using IMS-MS. A - native mass spectrum; B - drift time profiles for selected charged states; ms - milliseconds. Data acquired and analysed by Jun Yan.

Given that phosphorylation of the receiver domain of ExsG appeared to inactivate the HK core, it seemed likely that (Asp∼P)-ExsG adopts a conformational state different to that of unphosphorylated protein. Nucleotide binding to the ATP-binding

domain could also affect the shape of the protein; in the structures of canonical HK cores the ATP lid becomes ordered in the nucleotide-bound state of the CA domain. Even though in the case of ExsG the ATP lid may be shorter or absent (see Chapter 4), the binding of the nucleotide may still influence the structure of the HK core. Therefore the drift time profile of apo-ExsG was compared against the profiles representing (Asp∼P)-ExsG and nucleotide-bound ExsG to determine if the relative proportions of the two conformational states would change.

Notably, phosphorylation of the HK core of ExsG could also result in considerable conformational alterations affecting the collisional cross-section. However, initial attempts to study (His∼P)-ExsG failed as the proportion of phosphorylated molecules during the autokinase reaction was too low to be detected by mass spectrometry. This could be due to the equilibrium of the autophosphorylation reaction favouring the unphosphorylated protein; unfortunately, generating a sufficiently large proportion of ExsG molecules phosphorylated upon histidine to be detected by this method seemed unattainable.

6.3.1 Elucidation of the conformational changes associated with nucleotide binding

Determination of the binding constant for ExsG-ATP interaction by means of isothermal titration calorimetry was unsuccessful as the protein precipitated upon stirring; however, without stirring still no binding was detected and further experiments were not attempted. According to the data compiled by Krell and co-workers [21] the dissociation constants for the HK-ATP binding is commonly around 100 - 200 μM. Based on the preliminary results from autokinase assays performed at varying ATP concentrations it seems that the Michaelis constant (K_M), and probably the dissociation constant, are likely to fall within that micromolar range (see Fig. 4.9). Hence to obtain an ExsG population with approximately half of the molecules bound to ATP, the concentration of the nucleotide should be at least 100 μM. Initial experiments involved protein that was incubated with 0.1 - 0.2 mM AMP-PNP, an ATP analogue, for ∼30 minutes.

However, even though prior to the experiment AMP-PNP was purified by ethanol precipitation, the mass spectra indicated that a significant amount of adducts were associated with the protein. These adducts could represent remaining salts and contaminants from the AMP-PNP stock solution, or non-specifically bound additional nucleotides. Passing the ExsG-ATP incubation mixture through a chromatography spin column to remove the excess of the nucleotide and potential contaminants significantly

improved the quality of the spectra. An increase in the number of charges per protein ion was observed, suggesting that the protein was bound to AMP-PNP. The drift time profiles showed a decrease in intensity of the more compact species (drift time ∼10 ms), which implied that nucleotide binding stabilises the less compact form of ExsG.

Unfortunately, re-analysis of the data indicated that the observed charge shift did not correlate with a detectable mass shift. At the time when this thesis was written new data were acquired, using protein that was incubated with 1 mM AMP-PNP and 2 mM magnesium chloride for 30 min and subsequently purified using chromatography spin column. Even though the spectra again indicated the presence of adducts, mass shifts corresponding to bound protein species were evident. Data analysis was still in progress when this thesis was submitted, hence further details could not be provided here. However, preliminary analysis confirmed the earlier observations, which implied that the less compact conformation was stabilised by the binding of the nucleotide. This finding suggests that the ability to bind ATP may be a determining factor directing autokinase activity, as only one of the two conformational states allows binding of the nucleotide.

6.3.2 Elucidation of the conformational changes associated with phosphorylation of the receiver domain

Given that ExsG phosphoaspartate is stable over time, it was expected that (Asp∼P)-ExsG should be detected using IMS-MS. Initially, in order to ensure that ExsG receiver domain retained the ability to autophosphorylate in ammonium acetate buffer, a sample was analysed by phosphoprotein affinity gel electrophoresis; after 2 hours of incubation with PA at least half of the protein molecules were located within the upper band, confirming that phosphorylation from PA was efficient (Fig. 6.13). The protein was then incubated with PA for approximately 3 hours, after which it was subjected to two rounds of purification using chromatography spin columns and analysed by IMS-MS.

The obtained mass spectrum exhibited a shift of the mass and the charged states, suggesting that an additional charged group was present on the protein surface (Fig. 6.14A). Similarly to the apo-protein and protein bound to AMP-PNP, the two sub-populations of ExsG were only observed in the IMS-MS analysis in the case of ions with more associated charges (Fig. 6.14B). It appeared that after the prolonged incubation with PA the intensity of the form with a smaller collision cross-section (drift time ∼10 ms) was considerably higher, indicating that phosphorylation of the receiver domain may stabilise the more compact conformation of ExsG.

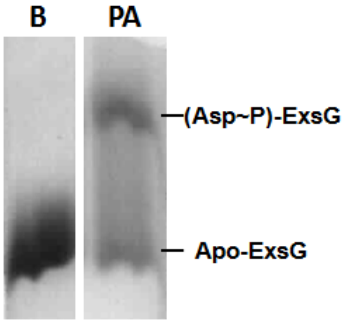


Figure 6.13: Test for ExsG receiver domain phosphorylation from PA in ammonium acetate buffer. B - before, PA - after 2 hours long incubation with PA

At the time when this thesis was written, a new data set was acquired and data analysis was not finished by the submission date. However, the trend was the same - phosphorylation of the receiver domain seemed to correlate with an increased proportion of the more compact form of ExsG. Furthermore, mass changes resulting from phosphorylation were detected in the mass spectrum, confirming that at least some of the subunits of the hexameric assembly were phosphorylated upon aspartate.

The observation that phosphorylation of the receiver domain correlates with an increase in the proportion of the more compact form of ExsG, while the binding of the nucleotide to the HK core stabilises the other conformation, suggests that the two conformations dynamically sampled by the unphosphorylated ExsG may be representing an active and inactive form of the HK core. Nucleotide binding and aspartate phosphorylation can shift this equilibrium in opposite directions. The less compact state, which allows the binding of ATP, was termed “open”, while the conformation that appears more stable in the IMS-MS analysis was termed “closed”. Considering that the results from both beryllofluoride and PA assays, presented earlier in this chapter, indicated that formation of the phosphoaspartate is associated with reduced autokinase activity, it seemed likely that in the “closed”, inactive conformation nucleotide binding to the ATP-binding domain would be hindered.

6.4 Summary

The RR activity of ExsG was not abolished upon phosphorylation of the HK core, while the autokinase activity was evidently downregulated when the receiver domain was phosphorylated. Therefore, in the context of the signalling cascade downstream of BphP1, ExsG acts as a RR with autokinase activity as the output regulated in a phosphorylation-dependent manner. The possibility that the presence of the phos-

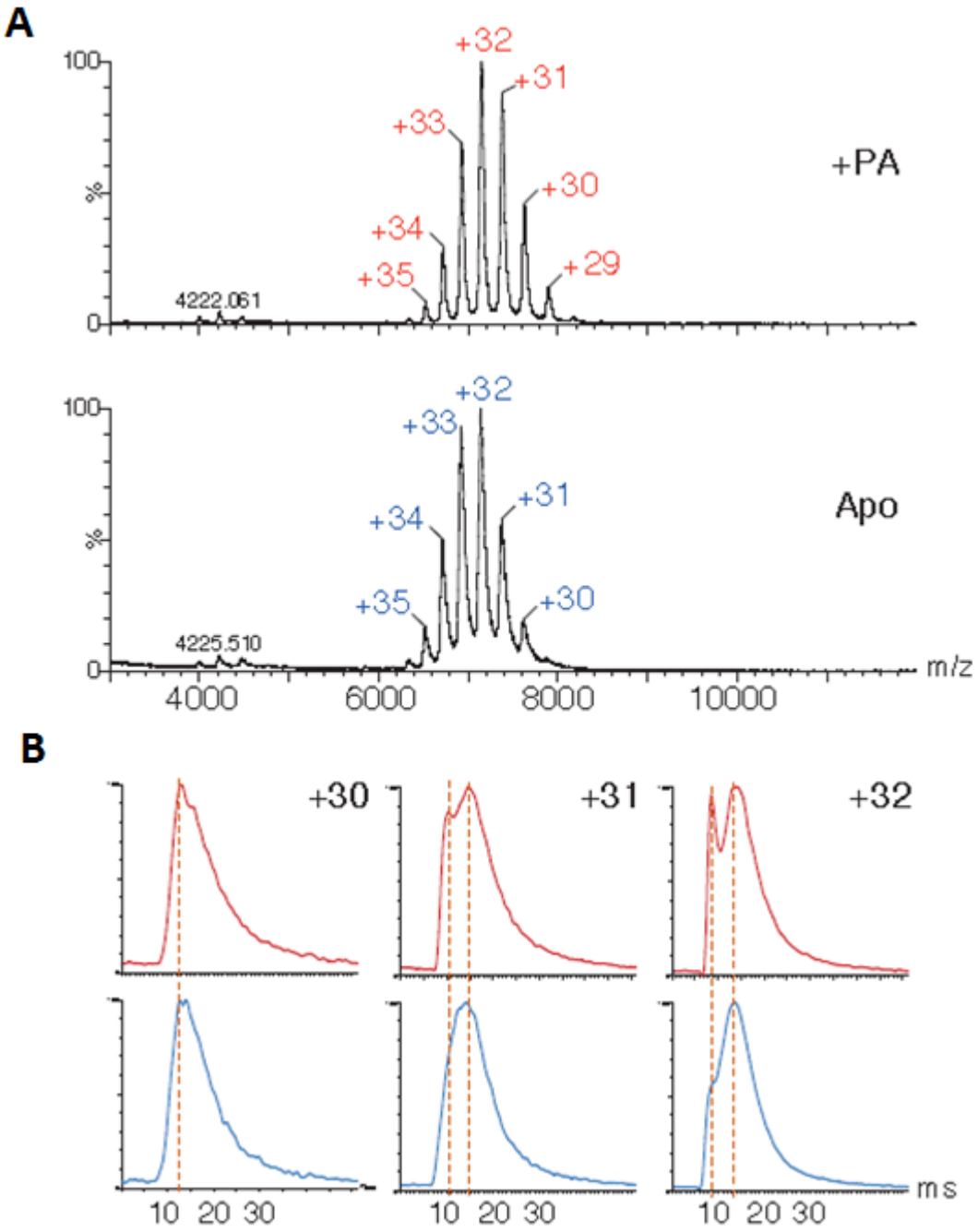


Figure 6.14: Conformational changes resulting from phosphorylation of the receiver domain of ExsG. A - mass spectra of unbound (apo) and phosphorylated (+PA) ExsG; B - drift time profiles for selected charged states. Red denotes protein phosphorylated by PA, blue - unbound protein (Apo). Dashed lines indicate the approximate peak positions of the two populations of molecules present. Data acquired and analysed by Jun Yan.

phoaspartate may affect the stability of the phosphohistidine was ruled out.

Preliminary IMS-MS results indicated that apo-ExsG exists in an equilibrium between two conformational states, characterised by differential stabilities at higher charge densities. Whereas phosphorylation of the receiver domain stabilises the more compact ("closed") form, nucleotide binding seems to exert the opposite effect. It appears likely that the "open" state allows nucleotide binding, while the "closed" form would represent the conformation precluding the binding of ATP, effectively inactivating the HK core.

Chapter 7

Conclusions and discussion

As discussed in the Introduction, inherent modularity and the enormous diversity of signals that can be perceived by the cell using TCSs make this form of signal transduction a promising addition to the field of synthetic biology. However, to exploit the full potential of two-component signal transduction it is crucial, first of all, to enrich the pool of available sensing, regulatory and phosphotransfer modules by extensive structural and functional characterisation of diverse protein domains involved in signalling. Furthermore, the mechanisms of interdomain signal transmission need to be fully elucidated to allow the design of functional chimaeras. Finally, it would be beneficial to determine the function of each TCS in its original and new environment, including the identification of stimuli sensed and responses triggered, and the investigation of cross-reactivity between signalling pathways in a cell of a model organism.

One of the aims of this project was to contribute to the field of two-component signal transduction, by *in silico* exploration of the *Rhizobium* NT-26 two-component signalome and identification of systems and proteins that may be interesting in the context of synthetic biology. The work described in this thesis also aimed to advance the general understanding of the physiological role of and the requirements for TCSs in controlling the responses to changes in the environment. Finally, the aim was to couple these analyses to characterisation of structural and biochemical properties of TCS proteins. Nearly 30 expression constructs that were generated in this work can be used for further mechanistic, structural and functional characterisation. Extensive *in vitro* studies of a complex TCS variant involving a bacteriophytochrome photoreceptor revealed a new type of branched pathway that utilises an novel signalling protein as an additional regulatory step. This hybrid RR, exhibiting both HK and RR activities, constitutes a novel type of building block that can be exploited in engineering intricate signalling circuits.

7.1 Diversity of TCSs in *Rhizobium* NT-26

The initial objective of this project was to characterise the two-component "signalome" of *Rhizobium* NT-26 in the context of other bacterial species; the results are presented in Chapter 3. Altogether 90 proteins were assigned as HKs or RRs based on the presence of the conserved domains involved in phosphotransfer. NT-26 "signalome", including the two-component "signalome" as well as all the other signalling proteins identified (chemoreceptors, Ser/Thr/Tyr protein kinases etc.), comprises 126 proteins. The calculated bacterial IQ (95) was comparable to the intelligence level of other rhizobia, although it could be described as average amongst other bacterial species, showing that free-living microbes with several metabolic strategies do not necessarily boast a considerable signalling complement. Notably, none of the putative 22 TCSs were unique, except for one canonical system whose homologues were identified in only one close rhizobial relative; the function of that TCS could not be assigned.

NT-26 is a bacterial "extrovert" as it monitors primarily extracellular parameters, which is evidenced by a significant proportion of proteins sensing extracytoplasmic stimuli. Most NT-26 TCSs seem to regulate gene expression, indicating that a number of stimuli may be coupled to activation of metabolic pathways, as exemplified by the AioS-AioR system, which has previously been shown to regulate the expression of arsenite oxidase in response to arsenite [112, 162]. The majority of NT-26 sensor HKs are membrane-bound and the most prominent class comprised periplasmic-sensing proteins. Aside from AioS and FeuQ none of the remaining 12 (13 including HK3041) periplasmic sensors could be assigned a putative function. Only in the case of two proteins the portions predicted to extrude into the periplasm were recognised as domains by SMART or Pfam; however, the identified putative ligand-binding domains - CHASE3 and 2CSK_N - have not been functionally or structurally characterised so far in any protein system or organism. Out of the five NT-26 proteins predicted to contain multiple transmembrane helices, one was assigned as NtrY and another as ActS, while the rest could not be functionally annotated. The current knowledge about extracytoplasmic and membrane-sensing mechanisms is therefore very limited. Given that NT-26 may respond to various chemicals and metals, some of these periplasmic or membrane-integral sensors could potentially be exploited, for example, in biosensor engineering, provided that their functions and structural properties are elucidated.

Among the 40 HKs possessing a sensory region, at least 12 contained two or more recognisable sensing domains, suggesting that they may monitor several parameters simultaneously. Out of these, 10 were either fully soluble or membrane anchored but

otherwise likely to constitute cytoplasmic sensors. This finding seems to support the hypothesis, proposed by Krell and co-workers [26], that the advantage of cytoplasmic HKs over one-component systems is the possibility of multi-sensing, which would account for the emergence of fully soluble TCSs. Even though classified as an "extrovert", NT-26 boasts a considerable complement of cytoplasmic sensor HKs, most of which exhibit different numbers and combinations of PAS and GAF domains. Notably, however, despite cytoplasmic localisation some of these proteins monitor extracellular stimuli, e.g. NtrB and BphPs.

The HWE type of HK core is relatively common in rhizobia, which has already been observed by Karniol and Vierstra [60]. NT-26 encodes 8 proteins predicted to contain the HWE_HK domain, two of which represent the distinct class of proteins termed "hybrid RRs" by Wuichet and co-workers [16]. Using ExsG as a case study, the structure and function of hybrid RRs are addressed later in this chapter. Considering the fact that the ATP-binding domains (ABDs) of HWE HKs are not usually annotated as canonical HATPase_c domains by SMART and Pfam, and yet ExsG ABD was classified into the GHKL superfamily, it is likely that they have diverged from CA domains a long time ago. Homology between the two HK families is further supported by the observation that HWE HK cores are topologically analogous to the canonical ones, and may comprise two distinct regions equivalent to DHp and CA domains. The boundaries of the HWE_HK domain defined by SMART and Pfam may therefore need to be revised as the current domain annotation includes the first two structural elements (a helix and a strand) of the putative ABD/CA domain (see Section 5.4).

Another interesting feature of NT-26 two-component "signalome" is the significant disproportion between the number of hybrid HKs (12) and histidine phosphotransfer domains (1, excluding CheA Hpt domain). Identification of Hpt domains is not straightforward due to low overall sequence conservation [69, 4]; however, even if the NT-26 genome encodes more proteins containing Hpt domains, it is unlikely that as many as 11 were missed in the extensive searches conducted herein. Under-representation of Hpt domains has been previously reported [134, 135] and linked to the need for convergent signalling, whereby the Hpt domain constitutes the focal point receiving signals from several HKs, or relaying the signal to multiple RRs. The responses triggered by different TCSs frequently overlap [17, 14], and the cell can economise on protein synthesis by introducing convergence points [135]. In a follow-up of this project, this hypothesis is going to be tested by comparing the abundance of hybrid HKs and Hpt domains across diverse species. In addition, NT-26 phosphorelay pathways appear to exhibit a high level of complexity; three hybrid HKs possessed more than one receiver domain, while one

hybrid RR contained a canonical HK core followed by a second receiver domain. Given that all of these receiver domains seemed likely to be functional, based on the presence of conserved active site residues, these multi-hybrids constitute interesting targets for functional characterisation in the context of two-component signalling mechanisms.

Finally, it is worth noting that the number of HKs and RRs encoded by NT-26 genome was revised after ExsG, initially classified as a HK, was demonstrated to act as a RR. Assuming the same functionality in the case of the other two hybrid RRs, in NT-26 there are 43 HKs and 47 RRs.

7.2 Signalling cascade downstream of BphP1

The presence of a light sensor in a bacterium isolated from a mine was intriguing. Nijmona and Lamparter [108] postulated that bacteriophytochromes may also be involved in monitoring temperature changes; the type of stimulus detected by NT-26 BphPs remains to be elucidated. However, based on homology to *Agrobacterium tumefaciens* BphPs it can be assumed that autokinase activity of NT-26 BphP1 would be associated with the ground state, but would be abolished upon conversion to the excited state.

As described in Chapter 4, NT-26 BphP1 and ExsG possess autokinase activity and the four proteins belonging to BphP1 gene cluster were demonstrated to form a branched TCS, with the signalling direction analogous to the gene arrangement. Interestingly, ExsF is encoded in the opposite direction to the other three proteins, which may suggest that its expression is regulated separately. The two single-domain RRs are likely to be the final steps within the pathway, potentially mediating protein-protein interactions with downstream targets.

Even though replacement of the HWE-motif histidine with a lysine in *A. tumefaciens* BphP2 abolished autokinase activity [60], ExsG is active despite the absence of the conserved histidine residue. This could be due to the structurally non-conservative nature of the substitution applied by Karniol and Vierstra. Other amino acids, for instance asparagine, should be tested at that position in HWE HKs to ascertain the significance of the HWE-motif histidine.

Autokinase assays indicated that BphP1 is more active than ExsG and the stand-alone HK core of ExsG, which could be a consequence of differences between canonical and HWE HKs in terms of ATP affinity. However, the preliminary results presented in Section 4.3.1.2 showed that the relationship between the phosphorylation extent of BphP1-HK and ATP concentration up to 1 mM is linear rather than hyperbolic. Given that the reaction rate depends directly on nucleotide concentration, the affinity

of BphP1-HK for ATP appears to be low. In contrast, ExsG autophosphorylation rate seems to be close to maximum in the presence of 1 mM ATP; as the value of Michaelis constant (K_M) is significantly lower than in the case of BphP1-HK autokinase reaction, ExsG may exhibit a higher affinity for ATP than BphP1. The considerable discrepancy between the activity levels is therefore likely to result from different rates of phosphotransfer between the two types of HK core. Notably, however, the activity of the full-length BphP1 would need to be tested, as the high autophosphorylation rate may result from the absence of the entire N-terminal portion.

In addition to autokinase activity, ExsG was herein shown to possess RR activity. Preliminary data indicated that BphP1 preferentially phosphorylates ExsG over AgR, but the rates of phosphotransfer did not appear dissimilar based on the results from separate transphosphorylation reactions; this implies that BphP1 may exhibit a higher affinity for ExsG. Given that phosphorylation of the ExsG receiver domain was shown to downregulate the autokinase activity (see Chapter 6), the emergence of phospho-BphP1 would result in phosphorylation of ExsG, but phosphotransfer from ExsG HK core onto ExsF would cease. AgR would also be phosphorylated by BphP1, but at a slower rate. The two single-domain RRs are hence unlikely to co-exist in the phosphorylated state. However, it is not known whether phosphorylation of either of these proteins constitutes an activation or an inactivation mechanism.

As discussed in the Introduction (Section 1.4.1), the responses triggered by bacteriophytochrome signalling remain largely unknown. Several computational and experimental strategies were therefore undertaken to identify the type of response triggered by the two single-domain RRs. Based on the temporal stability of phosphoresidues it was concluded that BphP1-initiated signalling cascade most probably affects gene expression. However, the absence of phosphatase activity reported herein does not rule out the involvement of negative regulation - phosphatase activity may be confined to proteins encoded elsewhere in the genome, and could significantly reduce the stability of phosphoaspartates. Even though the test for phosphotransfer onto the Hpt domain protein showed a negative result, it could be accounted for by the poor solubility of the Hpt protein, and further experimental evidence would be necessary to exclude the possibility that AgR and/or ExsF participate in a large-scale phosphorelay. Comparison to well-characterised single-domain RRs by means of multiple sequence alignments was inconclusive, primarily due to the fact that both proteins exhibit several unique sequence and structure motifs.

AgR and ExsF were subjected to extensive *in silico* characterisation and they were herein postulated to represent two distinct sub-classes of single-domain RRs. The

presence of several distinct sequence and structure motifs in RRs of AgR/Rcp type, that are phosphorylated by prokaryotic phytochromes, was previously reported based on the crystal structures of three Rcp proteins [56, 57]; in fact, Benda and co-workers suggested that these proteins constitute a subfamily of CheY-like RRs [57]. Considering that AgR/Rcp RRs exhibit a conserved extended loop motif and dimerise via the α4/β5/α5 interface, implicated in mediating protein-protein interactions, they could interact with unique downstream protein targets. Similarly, based on the NMR stucture of an ExsF homologue, Sma0114, Sheftic and co-workers postulated that RRs coupled to HWE HKs are structurally, and perhaps mechanistically, different from canonical CheY-type RRs. The absence of several structural features within ExsF sequence, particularly the Y-T coupling tyrosine, suggests that activation of ExsF-type RRs may involve a mechanism distinct to that in canonical RRs.

In order to identify the response triggered by AgR and ExsF, their potential interaction with NT-26 FliM would need to be addressed experimentally. The genomes of NT-26 and all bacteria that possess BphP genes could be thoroughly scanned for Hpt domains, and identified proteins subjected to phosphotransfer tests. Moreover, it is important to gain further structural and functional information about the two single-domain RRs, including the relationship between phosphorylation, activation and oligomeric state. Provided that the stimulus detected by NT-26 BphP1 is established, *in vivo* studies can be performed; each gene encoding one of the four proteins can be knocked out and the resulting phenotype examined in the context of motility and changes in gene expression patterns. Finally, computational methods can be applied to identify the surfaces on AgR and ExsF that engage in protein-protein interactions, and thus the potential downstream protein targets may be determined; this work is currently ongoing.

7.3 ExsG: a novel type of two-component signalling unit

As discussed in the Introduction (Section 1.3.2), phosphorelays are relatively common in microorganisms, and branched pathways have been previously reported. Here a distinct variant of a branched TCS, employing a novel form of a signalling unit, is described. Chapter 5 presents the results obtained from several biophysical experiments, which indicate that ExsG adopts a hexameric state, in contrast to canonical dimeric HKs. However, according to several lines of investigation, the HWE HK core of ExsG shares

homology with canonical cores. The inhibitory effect of receiver domain phosphorylation on ExsG autokinase activity is described in Chapter 6. The involvement of ExsG as a negative signalling mediator between BphP1 and ExsF may reflect the need for additional phosphorylation checkpoints - ExsG or ExsF could be targeted by phosphatases or phosphorylated by other HKs. Given that the two NT-26 BphPs may exhibit opposing roles, as their *A. tumefaciens* homologues [71], and thus trigger different responses, it is tempting to speculate that BphP2 may also participate in regulating ExsG and/or ExsF phosphorylation levels.

Based on the preliminary IMS-MS data, the homohexameric assemblies of ExsG appear to exist in an equilibrium between a more compact ("closed") and a less compact ("open") state. The latter conformation seems to be stabilised by ATP binding, while phosphorylation of the receiver domain of ExsG was demonstrated to stabilise the "closed" conformation. Considering the fact that Asp62 phosphorylation had an inhibitory effect on the autokinase activity, the "closed" state might represent an ExsG form that cannot autophosphorylate. On the other hand, the less compact conformation may correspond to the active state of the kinase core. As the interaction with the nucleotide stabilised the "open" form, it is conceivable that downregulation of the autokinase activity involves occlusion of the ATP binding site.

This equilibrium concept, in the context of the entire BphP1-initiated signalling cascade, is schematically depicted in Figure 7.1. In the diagram, BphP1 is shown as a dimer with sensing domains twisted around each other based on the structure of *Deinococcus radiodurans* BphP [109], while AgR and ExsF are represented as monomers according to the size-exclusion chromatography results, which indicated they are likely to be monomeric (see Section 5.2.2). Given that the high-resolution structures of AgR homologues indicate that AgR/Rcp RRs exist as dimers or homodimerise upon phosphorylation, it is likely that AgR also homodimerises in a phosphorylation-dependent manner, and this was indicated in the figure. Note that in the proposed model, ExsG molecules in the "closed" conformation cannot autophosphorylate, while phosphorylation of the receiver domain can occur irrespective of the conformational state. This is based on the finding that phosphorylation of the receiver domain downregulates autokinase activity, but phosphorylation of the HK core does not seem to abolish the RR activity.

In the absence of a high-resolution structure it is impossible to predict the three-dimensional arrangement of ExsG domains. Further complication stems from the fact that the HK core promotes the formation of the hexameric assembly, but it remains unknown which of the two constituent domains is involved in forming inter-subunit

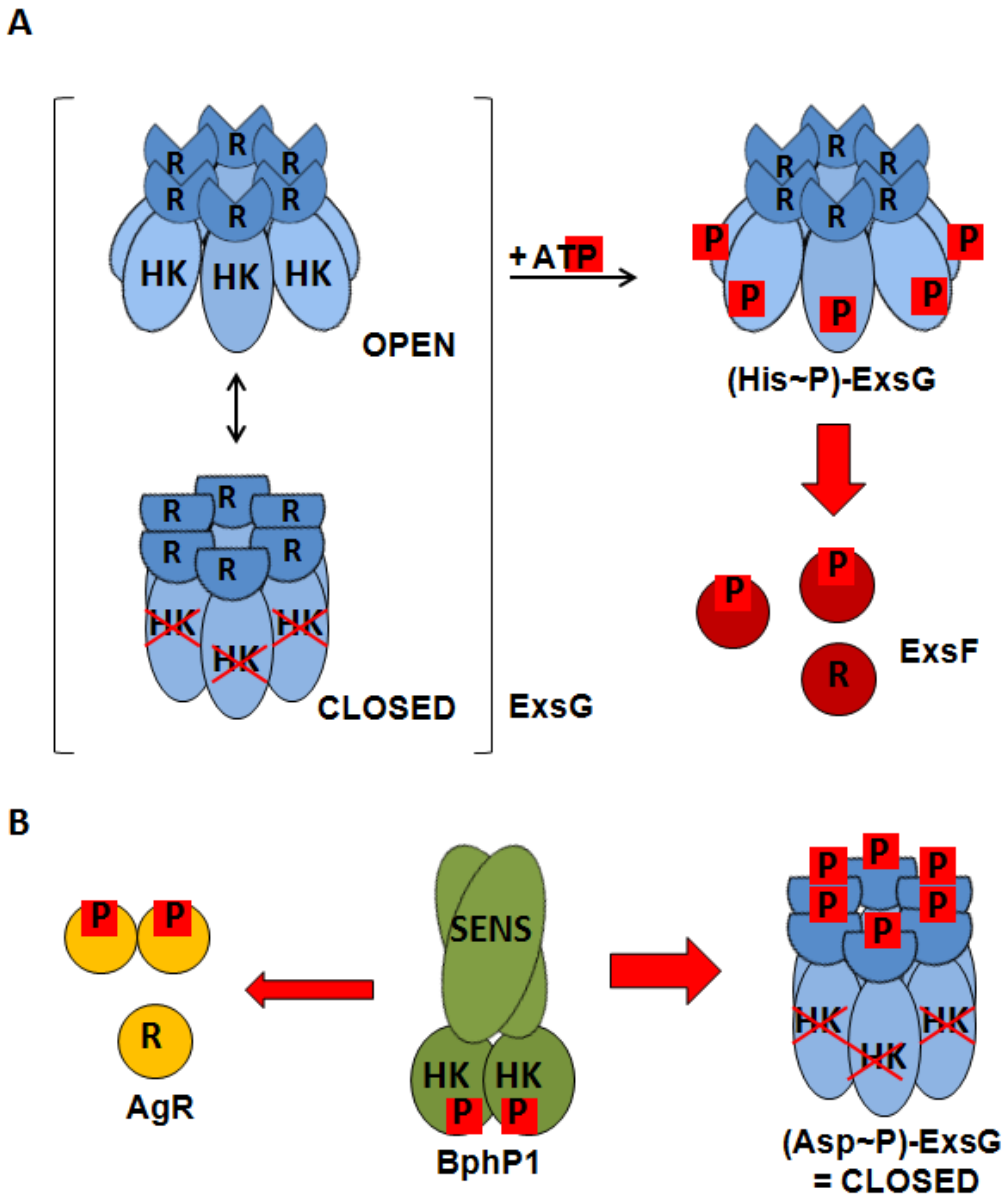


Figure 7.1: Schematic representation of the two-component signalling cascade downstream of BphP1. HK - HK core, R - receiver domain, SENS - sensing domain. Red arrows signify phosphotransfer reactions with different thicknesses in B indicating phosphotransfer preference. A - in the absence of phosphorylated BphP1, ExsG pre-exists in an equilibrium between "open" and "closed" states; the "open" conformation allows ATP binding and autophosphorylation and subsequent transfer of the phosphoryl group onto ExsF. B - when BphP1 autophosphorylates, the phosphoryl is transferred onto AgR and the receiver domain of ExsG. Conformational changes resulting from phosphorylation stabilise the "closed" state, downregulating autokinase activity. Phosphotransfer onto ExsF does not occur.

contacts. However, it was shown herein that the overall topology of ExsG HWE HK core resembles that of the canonical HK core, suggesting that the ATP-binding domain (ABD) and the histidine-containing domain (HCD) of ExsG may adopt the characteristic stem-and-leaf structure. Moreover, HCDs may dimerise to form a four-helix bundle - the notion that the functional unit of ExsG is a dimer was supported by the results from mass spectrometry. Importantly, based on the preliminary IMS-MS data, it appears that the autokinase activity may depend on the ability of the ABD domain to bind ATP, which further implies that the ABD could be in some way sequestered to preclude the binding of the nucleotide. Furthermore, phosphorylation of the receiver domain was shown to stabilise the more compact conformation of ExsG, indicating that additional or more extensive binding interfaces may have been formed - either between the consituent domains, or between the protomers.

In analogy to the model of regulation of canonical HK cores (see Section 1.1.2.1), structural changes within the phosphorylated receiver domain may affect the relative positioning of the HCD with respect to the ABD. Considering the fact that the extended interdomain linker probably leads from the receiver domain directly into the HCD, signal transmission may involve a piston-like movement or rotation of the helices. Alternatively, assuming that the receiver domain is positioned sufficiently close to the ABD, phosphorylation-induced conformational alterations within the receiver domain could promote, or disrupt, interdomain contacts with the ABD. Based on the assumption that the functional unit of ExsG is a homodimer, two models for ABD sequestration are proposed herein, schematically depicted in Figure 7.2. Notably, only *cis*-type of autophosphorylation was considered for simplicity, but the reaction could also occur in *trans*, between the ABD and the HCD located on different polypeptides.

Model A involves ABD sequestration by the receiver domain: in the "closed" conformation, stabilised by Asp62 phosphorylation, the position of the receiver domain would change, allowing the formation of interdomain contacts with the ABD. Formation of the interactions between the ABD and the receiver domain may also be accompanied by formation or disruption of ABD-HCD contacts. However, the observation that the stand-alone HK core of ExsG exhibits lower autokinase activity than the wild-type protein, at least under the conditions applied, suggests that this model may be insufficient to account for the inhibition of the HK core. Reduced activity of ExsG-HK might indicate that in the absence of the receiver domain the "closed", inactive state of the HK core is stabilised. This would imply that sequestration of the ABD cannot depend solely on the receiver domain, as then ExsG-HK would be more more active than the full-length protein. Therefore, a variation of the Model A is also proposed here, whe-

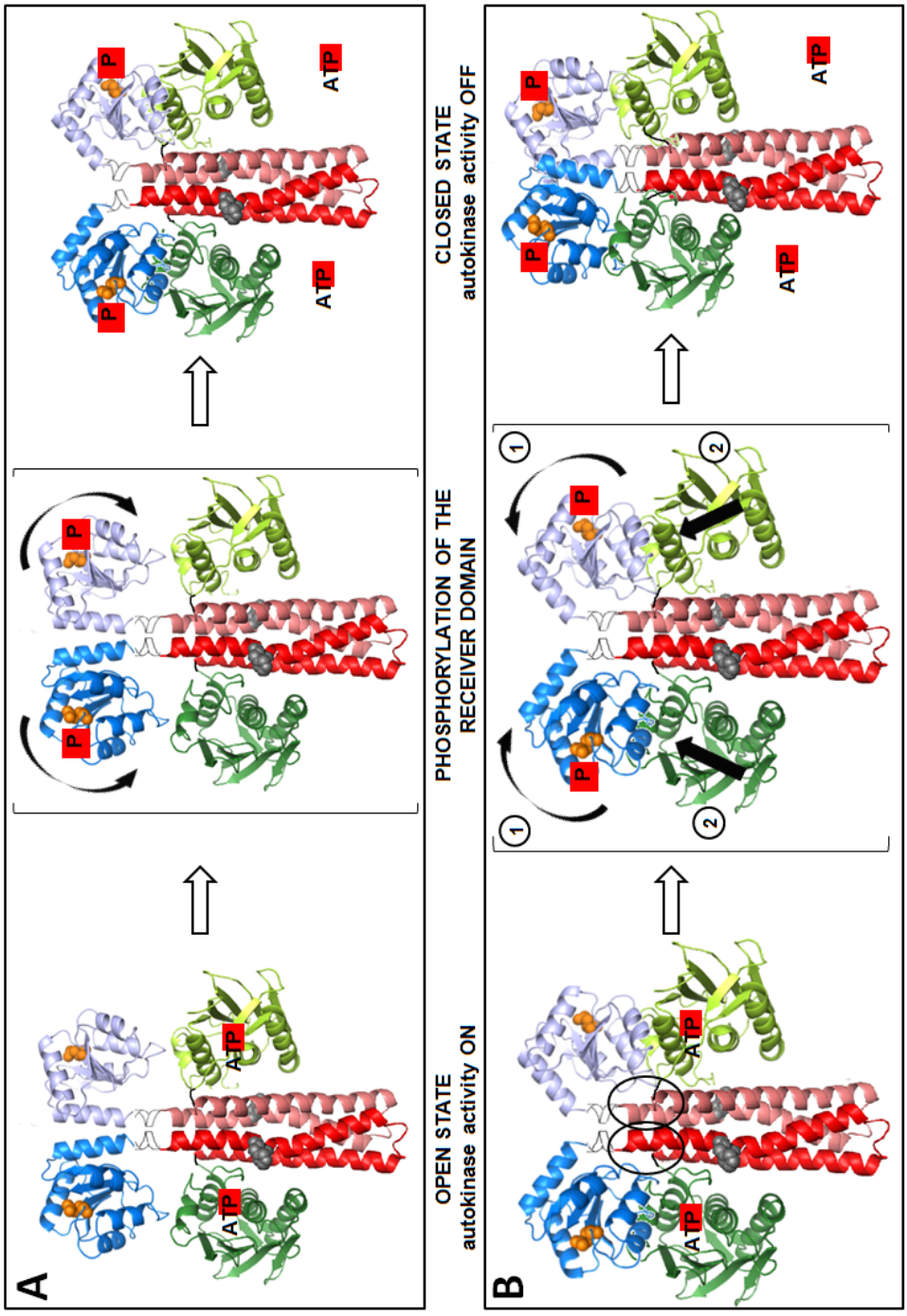


Figure 7.2: Proposed models of ExsG HK core inactivation by phosphorylation of the receiver domain. Ribbon structures of CheY (PDB 3CHY) and HK853 (3DGE) were used to represent separate ExsG domains. HCD is shown in red, ABD in green and receiver domain in blue. Phosphorylatable histidines (grey) and aspartates (orange) are shown as spheres. In model A sequestration of the ABD involves the formation of an interface between the phosphorylated receiver domain and the ABD. In model B, phosphorylated receiver domains dimerise (step 1) thus freeing the binding patch on the HCD (marked with black circles), and the ABD is sequestered by binding to the exposed binding patch (step 2).

reby interactions between the receiver domain and the ABD are present irrespective of the status of the receiver domain; in the "open" state, interdomain contacts would facilitate catalysis by positioning the ATP binding site close to the phosphorylatable histidine, but upon Asp62 phosphorylation a more extensive REC-ABD binding interface would be formed, occluding the ATP binding site. This possibility is not included in the schematic diagram.

Model B is analogous to the postulated model of CA domain sequestration by an exposed binding patch on the DHp domain (see Section 1.1.2.1), and implies that the ABD would be sequestered by binding to the HCD. Phosphorylation-induced exposure of the ABD-binding patch on the HCD surface could occur in two distinct ways - either by the movement of the HCD helices (e.g. "cogwheeling", see Section 1.1.2.1), or by displacement of the receiver domain. The latter option takes into account the possibility that both the ABD and the receiver domain could bind to the same surface on the HCD, but the receiver domain would bind with a higher affinity; upon phosphorylation, conformational changes within the receiver domain would somehow preclude the interaction between the receiver domain and the HCD, freeing the binding patch for the ABD. Notably, given that the stand-alone receiver domain of ExsG does not promote hexamerisation but may exhibit dimerisation propensity (see Section 5.2.2), the receiver domains within one dimeric unit may dimerise in a phosphorylation-dependent manner, which could effectively pull them away from the HCD.

The two models of autokinase activity inhibition are based on the assumption that the functional unit of ExsG is a dimer, whereas the hexameric state may in fact be necessary for interdomain signal transmission. In order to gain further insight into the function of ExsG-like hybrid RRs, it is necessary to establish the role of the higher-order structure; the hexameric assembly may imply cooperativity of, for example, ATP binding, or the involvement of allosteric regulation. First of all, it is crucial to determine if oligomerisation propensity is a common feature of other HWE HKs, such as BphP2. If the mode of oligomerisation is elucidated, it may be possible to disrupt the interfaces involved in the formation of the hexameric assembly, and verify if the resulting dimers retain RR and HK activity. Oligomerisation surfaces on ExsG may be identified by structural characterisation of isolated ExsG domains and computational methods. The positioning of the receiver domain with respect to the HCD could potentially be established by crystallisation of the two domains but excluding the ABD; crystallisation trials conducted in this project may have failed due to the presence of disordered regions, such as the predicted flexible linker between the HCD and the ABD. Furthermore, the impact of phosphorylation of one functional region on the activity of the other should

be assessed in quantitative terms. This would allow to determine the efficiency of HK core inhibition by phosphorylation of the receiver domain, and to establish if the RR activity of (His∼P)-ExsG molecules is affected.

Curiously, the analysis of the interdomain linker connecting the receiver domain with the HK core in several ExsG-like proteins revealed that canonical cores are preceded by a conserved structural element, the signalling helix, while HWE HK cores are not (see Section 5.4). In the case of ExsG and another NT-26 hybrid RR containing the HWE core a part of the interdomain linker exhibited a propensity to form a coiled-coil, but lacked several highly conserved sequence motifs and was significantly shorter than the signalling helix. This implies that, despite the overall similarity in terms of topology, the activity of HWE HK cores may be regulated in a different way to canonical HK cores - at least in the context of hybrid RRs. In the search for proteins containing the signalling helix, Anantharaman and co-workers identified 25 proteins with ExsG-like domain composition and arrangement [28], but it is not known whether this subset included any HWE HKs. It is also unknown if any of the other HKs found in their search possessed a HWE-type HK core. These sequence-related, structural and, potentially, mechanistic differences between classical and HWE-type HKs should be further investigated using a larger set of representative proteins.

Another interesting aspect in the context of HWE HKs is the mode of ATP binding. The distinct organisation of the ATP binding site, and the fact that the ATP lid is shorter or absent, may be in some way linked to the observation that ExsG autokinase activity depends on the ability to bind the nucleotide rather than on the exposure of the histidine. Differences in the mode of nucleotide binding may account for the lower activity of ExsG compared to BphP1. Importantly, however, it cannot be excluded that availability of the active site histidine also contributes to regulation of ExsG autokinase activity - this aspect has not been studied here. Equally, it remains to be tested whether canonical HKs exist in a similar equilibrium between “open” and “closed” states of the HK core. Further structural and functional studies are necessary to identify differences and similarities between canonical and HWE-type cores in terms of affinity for ATP, mechanisms of nucleotide binding, and the mode of signal transmission into the HK core from upstream domains. This information may help to determine if HWE HKs are indeed homologues of HisKA cores, and to elucidate any potential advantages of HWE-type cores in the context of two-component signal transduction.

7.4 Final remarks

The results from this project provide insight into the diversity of *Rhizobium* NT-26 two-component "signalome" and highlight the potential of two-component signal transduction in the context of synthetic biology applications. Despite the efforts, the response mediated by bacteriophytochrome photoreceptor 1 could not be identified, but the branched signalling pathway characterised herein constitutes a platform that can be incorporated into artificial signalling circuits. This basic scaffold can be "rewired" by changing the phosphotransfer specificities and substituting domains - for instance, engineering BphP1 to interact with RRs containing DNA-binding or enzymatic domains could generate a light-sensing pathway that evokes specific responses. Such modified pathways can be exploited *in vitro* as well as *in vivo*, producing microorganisms with programmable properties.

A novel type of signalling unit, exhibiting dual RR/HK activity and acting as a negative regulatory "switch", was reported here, enriching the pool of available two-component building blocks. Investigation of the mechanisms involved in coupling the phosphorylation of the N-terminal receiver domain with inactivation of the HK core is currently in progress. Further structural and functional studies are necessary to fully elucidate the role of ExsG-type proteins, and to verify if HWE HKs share homology with canonical HK cores. Importantly, the results presented in this thesis suggest that, despite the wealth of knowledge on two-component signal transduction, there are still many unexplored areas in this field.

Bibliography

[1] B. T. Nixon, C. W. Ronson, and F. M. Ausubel. Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *Proceedings of the National Academy of Sciences of the USA*, 83:7850–7854, 1986.

[2] J. S. Parkinson and E. C. Kofoid. Communication modules in bacterial signaling proteins. *Annual Review of Genetics*, 26:71–112, 1992.

[3] A. M. Stock, V. L. Robinson, and P. N. Goudreau. Two-component signal transduction. *Annual Review of Biochemistry*, 69:183–215, 2000.

[4] R. Gao and A. M. Stock. Biological insights from structures of two-component proteins. *Annual Review of Microbiology*, 63:133–154, 2009.

[5] Y. Gotoh, Y. Eguchi, T. Watanabe, S. Okamoto, A. Doi, and R. Utsumi. Two-component signal transduction as potential drug targets in pathogenic bacteria. *Current Opinion in Microbiology*, 13:232–239, 2010.

[6] R. Utsumi, R. E. Brissette, A. Rampersaud, S. A. Forst, K. Oosawa, and M. Inouye. Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate. *Science*, 245:1246–1249, 1989.

[7] A. Möglich, R. A. Ayers, and K. Moffat. Design and signaling mechanism of light-regulated histidine kinases. *Journal of Molecular Biology*, 385:1433–1444, 2009.

[8] A. J. Ninfa. Use of two-component signal transduction systems in the construction of synthetic genetic networks. *Current Opinion in Microbiology*, 13:240–245, 2010.

[9] A. Levskaya, A. A. Chevalier, J. J. Tabor, Z. Booth Simpson, L. A. Lavery, M. Levy, E. A. Davidson, A. Scouras, A. D. Ellington, E. M. Marcotte, and

C. A. Voigt. Synthetic biology: Engineering *Escherichia coli* to see light. *Nature*, 438:441–442, 2005.

[10] J. J. Tabor, H. M. Salis, Z. Booth Simpson, A. A. Chevalier, A. Levskaya, E. M. Marcotte, C. A. Voigt, and A. D. Ellington. A synthetic genetic edge detection program. *Cell*, 137:1272–1281, 2009.

[11] J. B. Stock, A J Ninfa, and A. M. Stock. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiological Reviews*, 53:450–490, 1989.

[12] V. L. Robinson, D. R. Buckler, and A. M. Stock. A tale of two components: a novel kinase and a regulatory switch. *Nature Structural Biology*, 7:626–633, 2000.

[13] R. C. Stewart. Protein histidine kinases: assembly of active sites and their regulation in signaling pathways. *Current Opinion in Microbiology*, 13:133–141, 2010.

[14] M. Goulian. Two-component signaling circuit structure and properties. *Current Opinion in Microbiology*, 13:184–189, 2010.

[15] E. D. Sonnenburg, J. L. Sonnenburg, J. K. Manchester, E. E. Hansen, H. C. Chiang, and J. I. Gordon. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing *in vivo* to carbohydrate metabolism. *Proceedings of the National Academy of Sciences of the USA*, 103:8834–8839, 2006.

[16] K. Wuichet, B. J. Cantwell, and I. B. Zhulin. Evolution and phyletic distribution of two-component signal transduction systems. *Current Opinion in Microbiology*, 13:219–225, 2010.

[17] M. T. Laub and M. Goulian. Specificity in two-component signal transduction pathways. *Annual Review of Genetics*, 41:121–145, 2007.

[18] P. M. Wolanin, P. A. Thomason, and J. B. Stock. Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biology*, 53:3013, 2002.

[19] T. Mascher, J. D. Helmann, and G. Unden. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiology and Molecular Biology Reviews*, 70:910–938, 2006.

[20] A. Khorchid and M. Ikura. Bacterial histidine kinase as signal sensor and transducer. *The International Journal of Biochemistry & Cell Biology*, 38:307–312, 2006.

[21] T. Krell, J. Lacal, A. Busch, H. Silva-Jiménez, M.-E. Guazzaroni, and J. L. Ramos. Bacterial sensor kinases: diversity in the recognition of environmental signals. *Annual Review of Microbiology*, 64:539–559, 2010.

[22] B.L. Taylor and I. B. Zhulin. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Review*, 63:479–506, 1999.

[23] M. Y. Galperin. Bacterial signal transduction network in a genomic perspective. *Environmental Microbiology*, 6:552–567, 2004.

[24] A. Möglich, R. A. Ayers, and K. Moffat. Structure and signalling mechanism of Per-ARNT-Sim domains. *Structure*, 17:1282–1294, 2009.

[25] J. Cheung and W. A. Hendrickson. Sensor domains of two-component regulatory systems. *Current Opinion in Microbiology*, 13:116–123, 2010.

[26] T. Krell, A. Busch, J. Lacal, H. Silva-Jiménez, and J.-L. Ramos. The enigma of cytosolic two-component systems: a hypothesis. *Environmental Microbiology Reports*, 1:171–176, 2009.

[27] M. Hulko, F. Berndt, M. Gruber, J. U. Linder, V. Truffault, A. Schultz, J. Martin, J. E. Schultz, A. N. Lupas, and M. Coles. The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, 126:929–940, 2006.

[28] V. Anantharaman, S. Balaji, and L. Aravind. The signaling helix: a common functional theme in diverse signaling proteins. *Biology Direct*, 1:25, 2006.

[29] A. E. Dago, A. Schug, A. Procaccini, J. A. Hoch, M. Weigt, and H. Szurmant. Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proceedings of the National Academy of Sciences of the USA*, 109:E1733–1742, 2012.

[30] M. B. Neiditch, M. J. Federle, A. J. Pompeani, R. C. Kelly, D. L. Swem, P. D. Jeffrey, B. L. Bassler, and F. M. Hughson. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell*, 6:1095–1108, 2006.

[31] H. Park, S. K. Saha, and M. Inouye. Two-domain reconstitution of a functional protein histidine kinase. *Proceedings of the National Academy of Sciences of the USA*, 95:6728–6732, 1998.

[32] P. Casino, V. Rubio, and A. Marina. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell*, 139:325–336, 2009.

[33] H. U. Ferris, S. Dunin-Horkawicz, N. Hornig, M. Hulko, J. Martin, J. E. Schultz, K. Zeth, A. N. Lupas, and M. Coles. Mechanism of regulation of receptor histidine kinases. *Structure*, 20:56–66, 2012.

[34] Y. Yang and M. Inoyue. Intermolecular complementation between two defective mutant signal-transducing receptors of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA*, 88:11057–11061, 1991.

[35] E. G. Ninfa, M. R. Atkinson, E. S. Kamberov, and A. J. Ninfa. Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. *Journal of Bacteriology*, 175:7024–7032, 1993.

[36] P. Casino, V. Rubio, and A. Marina. The mechanism of signal transduction by two-component systems. *Current Opinion in Structural Biology*, 20:763–771, 2010.

[37] S. Yamada, H. Sugimoto, M. Kobayashi, A. Ohno, H. Nakamura, and Y. Shiro. Structure of PAS-linked histidine kinase and the response regulator complex. *Structure*, 17:1333–1344, 2009.

[38] D. A. Jaques, D. B. Langley, C. M. Jeffries, K. A. Cunningham, W. F. Burkholder, J. M. Guss, and J. Trewhella. Histidine kinase regulation by a cyclophilin-like inhibitor. *Journal of Molecular Biology*, 384:422–435, 2008.

[39] M. J. Bick, V. Lamour, K. R. Rajashankar, Y. Gordiyenko, C. V. Robinson, and S. A. Darst. How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *Journal of Molecular Biology*, 386:163–177, 2009.

[40] D. Albanesi, M. Martin, F. Trajtenberg, M. C. Mansilla, A. Haouz, P. M. Alzari, D. de Mendoza, and A. Buschiazzo. Structural plasticity and catalysis regula-

tion of a thermosensor histidine kinase. *Proceedings of the National Academy of Sciences of the USA*, 106:16185–16190, 2009.

[41] J. F. Hess, R. B. Bourret, and M. I. Simon. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. *Nature*, 336:139–143, 1988.

[42] P. V. Attwood, M. J. Piggott, X. L. Zu, and P. G. Besant. Focus on phosphohistidine. *Amino Acids*, 32:145–156, 2007.

[43] R. B. Bourret. Receiver domain structure and function in response regulator proteins. *Current Opinion in Microbiology*, 13:142–149, 2010.

[44] S. A. Thomas, J. A. Brewster, and R. B. Bourret. Two variable active site residues modulate response regulator phosphoryl group stability. *Molecular Microbiology*, 69:453–465, 2008.

[45] Y. Pazy, A. C. Wollish, S. A. Thomas, P. J. Miller, E. J. Collins, R. B. Bourret, and R. E. Silversmith. Matching biochemical reaction kinetics to the timescales of life: Structural determinants that influence the autodephosphorylation rate of response regulator proteins. *Journal of Molecular Biology*, 392:1205–1220, 2009.

[46] E. Kinoshita, E. Kinoshita-Kikuta, K. Takiyama, and T. Koike. Phosphate-binding tag, a new toold to visualise phosphorylated proteins. *Molecular and Cellular Proteomics*, 5:749–757, 2006.

[47] L. J. Kenney. How important is the phosphatase activity of sensor kinases? *Current Opinion in Microbiology*, 13:168–176, 2010.

[48] A. Y. Mitrophanov and E. A. Groisman. Signal integration in bacterial two-component regulatory systems. *Genes & Development*, 22:2601–2611, 2008.

[49] T. N. Huynh and V. Stewart. Negative control in two-component signal transduction by transmitter phosphatase activity. *Molecular Microbiology*, 82:275–286, 2011.

[50] W.-S. Yeo, I. Zwir, H. V. Huang, D. Shin, A. Kato, and E. A. Groisman. Intrinsic negative feedback governs activation surge in two-component regulatory systems. *Molecular Cell*, 45:409–421, 2012.

[51] C. M. Dyer and F. W. Dahlquist. Switched or not?: the structure of unphosphorylated CheY bound to the N terminus of FliM. *Journal of Bacteriology*, 188:7354–7363, 2006.

[52] A. M. Stock and J. Guhaniyogi. A new perspective on response regulator activation. *Journal of Bacteriology*, 188:7328–7330, 2006.

[53] A. H. West and A. M. Stock. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences*, 26:369–375, 2001.

[54] U. Jenal and M. Y. Galperin. Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. *Current Opinion in Microbiology*, 12:152–160, 2009.

[55] R. Gao and A. M. Stock. Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Current Opinion in Microbiology*, 13:160–167, 2010.

[56] Y. J. Im, S.-H. Rho, C.-M. Park, S.-S. Yang, J.-G. Kang, J. Y. Lee, P.-S. Song, and S. H. Eom. Crystal structure of a cyanobacterial phytochrome response regulator. *Protein Science*, 11:614–624, 2002.

[57] C. Benda, C. Scheufler, N. Tandeau de Marsac, and W. Gärtner. Crystal structures of two cyanobacterial response regulators in apo- and phosphorylated form reveal a novel dimerization motif of phytochrome-associated response regulators. *Biophysical Journal*, 87:476–487, 2004.

[58] D. Higgins and J. Dworkin. Recent progress in *Bacillus subtilis* sporulation. *Journal of Molecular Biology*, 36:131–148, 2012.

[59] R. Dutta and M. Inouye. GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences*, 25:24–28, 2000.

[60] B. Karniol and R. D. Vierstra. The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling. *Journal of Bacteriology*, 186:445–453, 2004.

[61] C. Tomomori, T. Tanaka, R. Dutta, H. Park, S. K. Saha, Y. Zhu, R. Ishima, D. Liu, K. I. Tong, H. Kurokawa, H. Qian, M. Inouye, and M. Ikura. Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. *Nature Structural Biology*, 6:729–734, 1999.

[62] A. Marina, C. D. Waldburger, and W. A. Hendrickson. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *The EMBO Journal*, 24:4247–4259, 2005.

[63] T. Tanaka, S. K. Saha, C. Tomomori, R. Ishima, D. Liu, K. I. Tong, H. Park, R. Dutta, L. Qin, M. B. Swindells, T. Yamazaki, A. M. Ono, M. Kainosho, M. Inouye, and M. Ikura. NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature*, 396:88–92, 1998.

[64] J. Schultz, F. Milpetz, P. Bork, and C. P. Ponting. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences*, 95:5857–5864, 1998.

[65] R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, and A. Bateman. The Pfam protein families database. *Nucleic Acids Research*, 38:D211–D222, 2010.

[66] O. Ashenberg, K. Rozen-Gagnon, M. T. Laub, and A. E. Keating. Determinants of homodimerization specificity in histidine kinases. *Journal of Molecular Biology*, 413:222–235, 2011.

[67] J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian, and M. T. Laub. Rewiring the specificity of two-component signal transduction systems. *Cell*, 133:1043–1054, 2008.

[68] A. Marina, C. Mott, A. Auyzenberg, W. A. Hendrickson, and C. D. Waldburger. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. *Journal of Biological Chemistry*, 276:41182–41190, 2001.

[69] M. D. Baker, P. M. Wolanin, and J. B. Stock. Signal transduction in bacterial chemotaxis. *BioEssays*, 28:9–22, 2006.

[70] R. Dutta, L. Qin, and M. Inouye. Histidine kinases: diversity of domain organisation. *Molecular Microbiology*, 34:633–640, 1999.

[71] B. Karniol and R. D. Vierstra. The pair of bacteriophytochromes from *Agrobacterium tumefaciens* are histidine kinases with opposing photobiological properties. *Proceedings of the National Academy of Sciences of the USA*, 100:2807–2812, 2003.

[72] K. Volz and P. Matsumura. Crystal structure of *Escherichia coli* CheY refined at 1.7 å resolution. *Journal of Biological Chemistry*, 266:15511–15519, 1991.

[73] R. Gao, Y. Tao, and A. M. Stock. System-level mapping of *E. coli* response regulator dimerization with FRET hybrids. *Molecular Microbiology*, 69:1358–1372, 2008.

[74] R. B. Bourret. Signal transduction meets systems biology: deciphering specificity determinants for protein-protein interactions. *Molecular Microbiology*, 69:1336–1340, 2008.

[75] C. M. Barbieri, T. R. Mack, V. L. Robinson, M. T. Miller, and A. M. Stock. Regulation of response regulator autophosphorylation through interdomain contacts. *Journal of Biological Chemistry*, 285:32325–32335, 2010.

[76] L. E. Ulrich, E. V. Koonin, and I. B. Zhulin. One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology*, 13:52–56, 2005.

[77] E. J. Capra and M. T. Laub. Evolution of two-component signal transduction systems. *Annual Review of Microbiology*, 66:325–347, 2011.

[78] P. J. A. Cock and D. E. Whitworth. Evolution of prokaryotic two-component system signaling pathways: gene fusions and fissions. *Molecular Biology and Evolution*, 24:2355–2357, 2007.

[79] K. K. Koretke, A. N. Lupas, P. V. Warren, M. Rosenberg, and J. R. Brown. Evolution of two-component signal transduction. *Molecular Biology and Evolution*, 17:1956–1970, 2000.

[80] D. Kim and S. Forst. Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology*, 147:1197–1212, 2001.

[81] M. Galperin. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *BMC Microbiology*, 5:35, 2005.

[82] E. Alm, K. Huang, and A. Arkin. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Computational Biology*, 2:e143, 2006.

[83] M. Y. Galperin, A. N. Nikolskaya, and E. V. Koonin. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiology Letters*, 203:11–21, 2001.

[84] W. Qian, Z.-J. Han, and C. He. Two-component signal transduction systems of *Xanthomonas* spp.: a lesson from genomics. *PLoS Genetics*, 21:151–161, 2008.

[85] J. Lawrence. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Current Opinion in Genetics and Development*, 9:642–648, 1999.

[86] V. Raghavan and E. A. Groisman. Orphan and hybrid two-component system proteins in health and disease. *Current Opinion in Microbiology*, 13:226–231, 2010.

[87] S. Wegener-Feldbrügge and L. Sogaard-Andersen. The atypical hybrid histidine protein kinase RodK in *Myxococcus xanthus*: spatial proximity supersedes kinetic preference in phosphotransfer reactions. *Journal of Bacteriology*, 191:1765–1776, 2009.

[88] J. Zapf, U. Sen, Madhusudan, J. A. Hoch, and K. I. Varughese. A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure*, 8:851–862, 2000.

[89] J. M. Skerker, M. S. Prasol, B. S. Perchuk, E. G. Biondi, and M. T. Laub. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biology*, 3:e334, 2005.

[90] C. H. Bell, S. L. Porter, A. Strawson, D. I. Stuart, and J. P. Armitage. Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. *PLoS Biology*, 8:e1000306, 2010.

[91] E. J. Capra, B. S. Perchuk, E. A. Lubin, O. Ashenberg, J. M. Skerker, and M. T. Laub. Systematic dissection and trajetory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. *PLoS Genetics*, 6:e1001220, 2010.

[92] E. J. Capra, B. S. Perchuk, J. M. Skerker, and M. T. Laub. Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell*, 150:222–232, 2012.

[93] M. L. López-Redondo, F. Moronta, P. Salinas, J. Espinosa, R. Cantos, R. Dixon, A. Marina, and A. Contreras. Environmental control of phosphorylation pathways in a branched two-component system. *Molecular Microbiology*, 78:475–489, 2010.

[94] H. Smith. Phytochromes and light signal perception by plants - an emerging synthesis. *Nature*, 407:585–591, 2000.

[95] S. Yoshihara, F. Suzuki, H. Fujita, X. X. Geng, and M. Ikeuchi. Novel putative photoreceptor and regulatory genes required for the positive phototactic movement of the unicellular motile cyanobacterium *Synechocystis* sp. PCC 6803. *Plant and Cell Physiology*, 41:1299–1304, 2000.

[96] E. Giraud, J. Fardoux, N. Fourrier, L. Hannibal, B. Genty, P. Bouyer, B. Dreyfus, and A. Verméglio. Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria. *Nature*, 417:202–205, 2002.

[97] E. Giraud, S. Zappa, L. Vuillet, J.-M. Adriano, L. Hannibal, J. Fardoux, C. Berthomieu, P. Bouyer, D. Pignol, and A. Vermeglio. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodopseudomonas palustris*. *Journal of Biological Chemistry*, 280:32389–32397, 2005.

[98] S. J. Davis, A. V. Vener, and R. D. Vierstra. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic Eubacteria. *Science*, 286:2517–2520, 2012.

[99] S.-H. Bhoo, S. J. Davis, J. Walker, B. Karniol, and R. D. Vierstra. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature*, 414:776–779, 2001.

[100] B. Karniol, J. R. Wagner, J. M. Walker, and R. D. Vierstra. Phylogenetic analysis of the phytochrome superfamily reveals distinct microbial subfamilies of photoreceptors. *Biochemical Journal*, 392:103–116, 2005.

[101] N. C. Rockwell, Y.-S. Su, and J. C. Lagarias. Phytochrome structure and signalling mechanisms. *Annual Review of Plant Biology*, 57:837–858, 2006.

[102] K. Yeh and J. C. Lagarias. Eukaryotic phytochromes: Light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proceedings of the National Academy of Sciences of the USA*, 95:13976–13981, 1998.

[103] K. Yeh, S.-H. Wu, J. T. Murphy, and J. C. Lagarias. A cyanobacterial phytochrome two-component light sensory system. *Science*, 277:1505–1508, 1997.

[104] H. R. Bonomi, D. M. Posadas, G. Paris, M. del Carmen Carrica, M. Frederickson, L. I. Pietrasanta, R. A. Bogomolni, A. Zorreguieta, and F. A. Goldbaum. Light regulates attachment, exopolysaccharide production, and nodulation in *Rhizobium leguminosarum* through a LOV-histidine kinase photoreceptor. *Proceedings of the National Academy of Sciences of the USA*, 109:12135–12140, 2012.

[105] X. Yang, J. Kuk, and K. Moffat. Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: Photoconversion and signal transduction. *Proceedings of the National Academy of Sciences of the USA*, 105:14715–14720, 2008.

[106] T. Lamparter, N. Michael, F. Mittman, and B. Esteban. Phytochrome from *Agrobacterium tumefaciens* has unusual spectral properties and reveals an N-terminal chromophore attachment site. *Proceedings of the National Academy of Sciences of the USA*, 99:11628–11633, 2002.

[107] R. Tasler, T. Moises, and N. Frankenberg-Dinkel. Biochemical and spectroscopic characterization of the bacterial phytochrome of *Pseudomonas aeruginosa*. *FEBS Journal*, 272:1927–1936, 2005.

[108] I. Njimona and T. Lamparter. Temperature effects on *Agrobacterium* phytochrome Agp1. *PLoS ONE*, 6:e25977, 2011.

[109] Hua Li, J. Zhang, R. D. Vierstra, and Huilin Li. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proceedings of the National Academy of Sciences of the USA*, 107:10872–10877, 2010.

[110] L. Vuillet, M. Kojadinovic, S. Zappa, M. Jaubert, J.-M. Adriano, J. Fardoux, L. Hannibal, D. Pignol, A. Verméglio, and E. Giraud. Evolution of a bacteriophytochrome from light to redox sensor. *The EMBO Journal*, 26:3322–3331, 2007.

[111] J. M. Santini, L. I. Sly, R. D. Schnagl, and J. M. Macy. A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Applied and Environmental Microbiology*, 66:92–97, 2000.

[112] S. Sardiwal, J. M. Santini, T. H. Osborne, and S. Djordjevic. Characterization of a two-component signal transduction system that controls arsenite oxidation in the chemolithoautotroph NT-26. *FEMS Microbiology Letters*, 313:20–28, 2010.

[113] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389–4302, 1997.

[114] The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 40:D71–D75, 2012.

[115] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32:1792–1797, 2004.

[116] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. Jalview version 2: A multiple sequence alignment and analysis workbench. *Bioinformatics*, 25:1189–1191, 2009.

[117] A. Lupas, M. Van Dyke, and J. Stock. Predicting coiled coils from protein sequences. *Science*, 252:1162–1164, 1991.

[118] D. W. A. Buchan, S. M. Ward, A. E. Lobley, T. C. O. Nugent, K. Bryson, D. T. Jones, D. W. Buchan, T. C. Nugent, K. Bryson, and D. T. Jones. Protein annotation and modelling servers at University College London. *Nucleic Acids Research*, 38:W563–W568, 2010.

[119] C. Cole, J. D. Barber, and G. J. Barton. The Jpred3 secondary structure prediction server. *Nucleic Acids Research*, 36:W197–W201, 2008.

[120] A. Lobley, M. I. Sadowski, and D. T. Jones. pGenTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics*, 25:1761–1767, 2009.

[121] D. W. Buchan, A. J. Shepherd, D. Lee, F. M. Pearl, S. C. Rison, J. M. Thornton, and D. T. Jones. Gene3D: structural assignment for whole genes and genomes using the cath domain structure database. *Genome Research*, 12:503–514, 2002.

[122] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.

[123] R. C. Sheridan, J. F. McCullough, and Z. T. Wakefield. Phosphoramidic acid and its salts. *Inorganic Syntheses*, 13:23–26, 1971.

[124] P. Schuck. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal*, 78:1606–1619, 2000.

[125] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Slade, K. Thalassinos, R. Bateman, M. T. Bowers, and J. H. Scrivens. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*, 261:1–12, 2007.

[126] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, and M. Mann. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, 1:2856–2860, 2006.

[127] S. J. Geromanos, J. P. Vissers, J. C. Silva, C. A. Dorschel, G. Z. Li, M. V. Gorenstein, R. H. Bateman, and J. L. Langridge. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependent LC-MS/MS. *Proteomics*, 9:1683–1695, 2009.

[128] Y. Zhu and M. Inouye. The HAMP linker in histidine kinase dimeric receptors is critical for symmetric transmembrane signal transduction. *Journal of Biological Chemistry*, 279(46):48152–48158, 2004.

[129] R. Heermann, A. Fohrmann, K. Altendorf, and K. Jung. The transmembrane domains of the sensor kinase KdpD of *Escherichia coli* are not essential for sensing K^+ limitation. *Molecular Microbiology*, 47:839–848, 2003.

[130] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. Predicting transmembrane protein topology with a Hidden Markov Model: Application to complete genomes. *Journal of Molecular Biology*, 305:567–580, 2001.

[131] B. P. O'Hara, R. A. Norman, P. T. C. Wan, S. M. Roe, T. E. Barrett, R. E. Drew, and L. H. Pearl. Crystal structure and induction mechanism of AmiC-AmiR: a ligand-regulated transcription antitermination complex. *EMBO Journal*, 18:5175–5186, 1999.

[132] M. Greck, J. Platzer, V. Sourjik, and R. Schmitt. Analysis of a chemotaxis operon in *Rhizobium meliloti*. *Molecular Microbiology*, 15:989–1000, 1995.

[133] A. Ferré, J. de la Mora, T. Ballado, L. Camarena, and G. Dreyfus. Biochemical study of multiple CheY response regulators of the chemotactic pathway of *Rhodobracter sphaeroides*. *Journal of Bacteriology*, 186:5172–5177, 2004.

[134] W. Zhang and L. Shi. Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. *Microbiology*, 151:2159–2173, 2005.

[135] J. L. Hsu, H. C. Chen, H. L. Peng, and H. Y. Chang. Characterization of the histidine-containing phosphotransfer protein B-mediated multistep phosphorelay system in *Pseudomonas aeruginosa* pao1. *Journal of Biological Chemistry*, 283:9933–9944, 2008.

[136] B. Muller and J. Sheen. Arabidopsis cytokinin signaling pathway. *Science's STKE*, 2007:1–5, 2007.

[137] R. P. Tiwari, W. G. Reeve, M. J. Dilworth, and A. R. Glenn. Acid tolerance in *Rhizobium meliloti* strain WSM419 involves a two-component sensor-regulator system. *Microbiology*, 142:1693–1704, 1996.

[138] B. J. Fenner, R. P. Tiwari, W. G. Reeve, M. J. Dilworth, and A. R. Glenn. *Sinorhizobium medicae* genes whose regulation involves the ActS and/or ActR signal transduction proteins. *FEMS Microbiology Letters*, 236:21–31, 2004.

[139] K. H. Yeoman, M.-J. Delgado, M. Wexler, J. A. Downie, and A. W. B. Johnston. High affinity iron acquisition in *Rhizobium leguminosarum* requires the *cycHJKL* operon and the *feuPQ* gene products, which belong to the family of two-component transcriptional regulators. *Microbiology*, 143:127–134, 1997.

[140] T. C. Charles and E. W. Nester. A chromosomally encoded two-component sensory transduction system is required for virulence of *Agrobacterium tumefaciens*. *Journal of Bacteriology*, 175:6614–6625, 1993.

[141] H.-P. Cheng and G. C. Walker. Succinoglycan production by *Rhizobium meliloti* is rgulated through the ExoS-ChvI two-component regulatory system. *Journal of Bacteriology*, 180:20–26, 1998.

[142] L. Wang, F. Ma, Y. Qu, D. Sun, J. Guo, and B. Yu. Characterisation of a compound bioflocculant produced by mixed culture of *Rhizobium radiobacter* F2 and *Bacillus sphaeicus* F6. *World Journal of Microbiology and Biotechnology*, 27:2559–2565, 2011.

[143] N. Labbe, S. Parent, and R. Villemur. *Nitratireductor aquibiodomus* gen. nov., sp. nov,. a novel alpha-proteobacterium from the marine denitrification system

of the Montreal Biodome (Canada). *International Journal of Systematic and Evolutionary Microbiology*, 54:269–273, 2004.

[144] C. R. Anderson, G. J. Dick, M-L. Chu, J-C. Cho, R. E. Davis, S. L. Bräuer, and B. M. Tebo. *Aurantimonas manganoxydans*, sp. nov., and *Aurantimonas litoralis*, sp. nov.: Mn(II) oxidizing representatives of a globally distributed clade of alpha-*Proteobacteria* from the order *Rhizobiales*. *Geomicrobiology Journal*, 26:189–198, 2009.

[145] M. Y. Galperin, R. Higdon, and E. Kolker. Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. *Molecular bioSystems*, 6:721–728, 2010.

[146] G. H. Wadhams and J. P. Armitage. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular Cell Biology*, 5:1024–1037, 2004.

[147] J. W. Willett and J. R. Kirby. Genetic and biochemical dissection of a HisKA domain identifies residues requied exclusively for kinase and phosphatase activities. *PLoS Genetics*, 8:e1003084, 2012.

[148] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. *The proteomics protocols handbook*, chapter Protein identification and analysis tools on the ExPASy server, pages 571–607. Humana Press, 2005.

[149] G. S. Lukat and J. B. Stock. Response regulation in bacterial chemotaxis. *Journal of Cellular Biochemistry*, 1993:41–46, 1993.

[150] S. L. Porter and J. P. Armitage. Phosphotransfer in *Rhodobacter sphaeroides* chemotaxis. *Journal of Molecular Biology*, 324:35–45, 2002.

[151] T. Hübschmann, H. J. M. M. Jorissen, T. Börner, W. Gärtner, and N. Tandeau de Marsac. Phosphorylation of proteins in the light-dependent signalling pathway of a filamentous cyanobacterium. *European Journal of Biochemistry*, 268:3383–3389, 2001.

[152] A. M. Stock, J. M. Mottonen, J. B. Stock, and C. E. Schutt. Three-dimensional structure of chey, the response regulator of bacterial chemotaxis. *Nature*, 337:745–749, 1989.

[153] S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, E. A. Berry, and D. E. Wemmer. Crystal structure of an activated CheY. *Journal of Biological Chemistry*, 276:16425–16431, 2001.

[154] S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer. Crystal structure of an activated response regulator bound to its target. *Nature Structural Biology*, 8:52–56, 2001.

[155] S. R. Sheftic, P. P. Garcia, V. L. Robinson, D. J. Gage, and A. T. Alexandrescu. NMR assignments for the *Sinorhizobium meliloti* response regulator Sma0114. *Biomolecular NMR Assignments*, 5:55–58, 2011.

[156] P. H. Brown and P. Schuck. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. *Computer Physics Communications*, 178:105–120, 2008.

[157] H. J. Sterling, J. D. Batchelor, D. E. Wemmer, and E. R. Williams. Effects of buffer loading for electrospray ionization mass spectrometry of a noncovalent protein complex that requires high concentrations of essential salts. *Journal of American Society for Mass Spectrometry*, 21:1045–1049, 2010.

[158] F. Trajtenberg, M. Grana, N. Ruetalo, H. Botti, and A. Buschiazzo. Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase. *Journal of Biological Chemistry*, 285:24892–24903, 2010.

[159] A. L. Cuff, I. Sillitoe, T. Lewis, A. B. Clegg, R. Rentzsch, N. Furnham, M. Pellegrini-Calace, D. Jones, J. Thornton, and C. A. Orengo. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. *Nucleic Acids Research*, 39:D420–D426, 2011.

[160] D. Yan, H. S. Cho, C. A. Hastings, M. M. Igo, S.-Y. Lee, J. G. Pelton, V. Stewart, D. E. Wemmer, and S. Kustu. Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proceedings of the National Academy of Sciences of the USA*, 96:14789–14794, 1999.

[161] E. Kinoshita and E. Kinoshita-Kikuta. Improved Phos-tag SDS-PAGE under neutral pH conditions for advanced protein phosphorylation profiling. *Proteomics*, 11:319–323, 2011.

[162] G. Liu, M. Liu, E. H. Kim, W. S. Maaty, B. Bothner, B. Lei, C. Rensing, G. Wang, and T. R. McDermott. A periplasmic arsenite-binding protein involved in regulating arsenite oxidation. *Environmental Microbiology*, 14:1624–1634, 2012.

Appendix A

Primers and expression vector

Gene ID	Forward primer	Reverse primer
1631	gc**tcATGA**ACAAGCCCCAGCC (64)	cag**aagctt**TCATTGGGCTTTTCCTGTC (65)
1632_{1-157}	gccc**ATGg**ttGCCACCATCCTTGTCG (70)	cag**aagctt**CTGCATGAGGCGGGCAAG (70) (*)
$0925_{186-470}$	ga**ttcatga**ctCGCAATGCGCTGAAGCGG (70)	cg**aagctt**TTAGGGCCTGAACGGCGG (70)
0924	ga**tcATGa**ctGCCCCCGCCCGCCACGCC (77)	ccg**aagctt**TCACGCCTCGTCCTTGATCATGTAGCC (74)
3236	gccc**ATGg**tgAACCCGACGAGAAGCG (71)	gc**aagctT**CATGTCCTTCCTGTCGGTTGC (70)
3237	gc**tcATGA**TGGAAGTTCTCATCGTCG (65)	cag**aagctt**TCAATTCTGCTTGACCCC (65)
$1656_{109-384}$	ga**tcatga**ctCGGATCGCCGCCCCCTTG (72)	ccg**aagctt**TCAGCTTCGGGCGCCG (71)
1657	gc**tcATGa**cttTTCTCTTCTCTGATTTTGCTTG (66)	ccg**aagctt**CTAGCCATTGTCGGATC (66)
$0917_{241-468}$	ga**ccatgg**ctCGCCGCATCGTCGACCGTG (73)	gcg**aagctt**TCAGGCAACATCCTTCGTC (65)
0916	gccc**ATGg**ctCGCATTCTCGTAGTT (65)	cgc**aagctt**TCAGCTGACGTCGGTCG (67)
$1633_{321-756}$	ga**ccatgG**CCGTTGCCGACCGGATCGTT (71)	gcc**aagctt**TCAGACGCCATATGCGAGT (67)
1634_{1-130}	gccc**ATGG**CGTCTGATATTCTGGTCGTG (66)	gtt**aagctt**TCAGCTGACCTCGCGCTTC (65)
$3039_{211-543}$	gt**ccatgg**ctCAAAAGGCTGCCCTGA (67)	gcc**aagctT**CATTGTGTGAATTCTGCTCC (64)
3038	gccc**ATGg**ctACGAACATCCGTGTCGC (69)	gcc**aagctt**TCAGATCGCGCGCGTCTG (69)
$3863_{301-594}$	gccc**atgg**ctTCCTCGACCATAGCCAAT (68)	gcc**aagctt**TCATGCAGGCTCTCCAGC (68)
3864	gccc**ATGg**ctCAGACAATCGCCCTGGTG (70)	gcc**aagctt**TCAGGTGGCTTCGCGGAAG (69)
$4001_{191-462}$	ga**ccatgg**ctCCGCTCTACCGCTTCAG (68)	gcc**aagctt**CTACTCGGCGTGATCCC (67)
4002	gg**ccatgG**TGCGAATTCTTCTGGTG (64)	ga**caagctt**TCATCCAGCGGTTGCC (64)
$2937_{321-614}$	ga**ccatgg**ctCAGCGCCGGGCGCGGCT (77)	gg**caagctT**CATTCCGCCGCCTTCAGGGC (72)
2936_{1-140}	ga**ccATGg**ctAGCGAGCAGCGCGTCTTTC (70)	ga**caagctt**TCAGTCGTCCCGCTTGCCGG (70)
0152	gg**ccATGG**ATTTGGCAGTCTCACC (64)	gc**aagctt**TCAGTCCAGTAGCTCGC (64)
0153	gg**ccATGg**ctAATCCGGATGTGAACAGCC (68)	gg**aagctt**CTACGACACCCGCCCAGTG (68)

Table A.1: PCR primers (5'-3') used for amplification of selected NT-26 genes or their portions (excluding those listed in Table 2.1). Numbers in subscript indicate residues incorporated in the construct. Coding sequence is in capital letters, restriction enzyme sites in bold, stop codons engineered into partial sequences are underlined. Numbers in brackets are average melting temperatures in °C. (*) denotes a reverse primer missing a stop codon; the resulting sequence was later subjected to site-directed mutagenesis to introduce the missing stop codon.

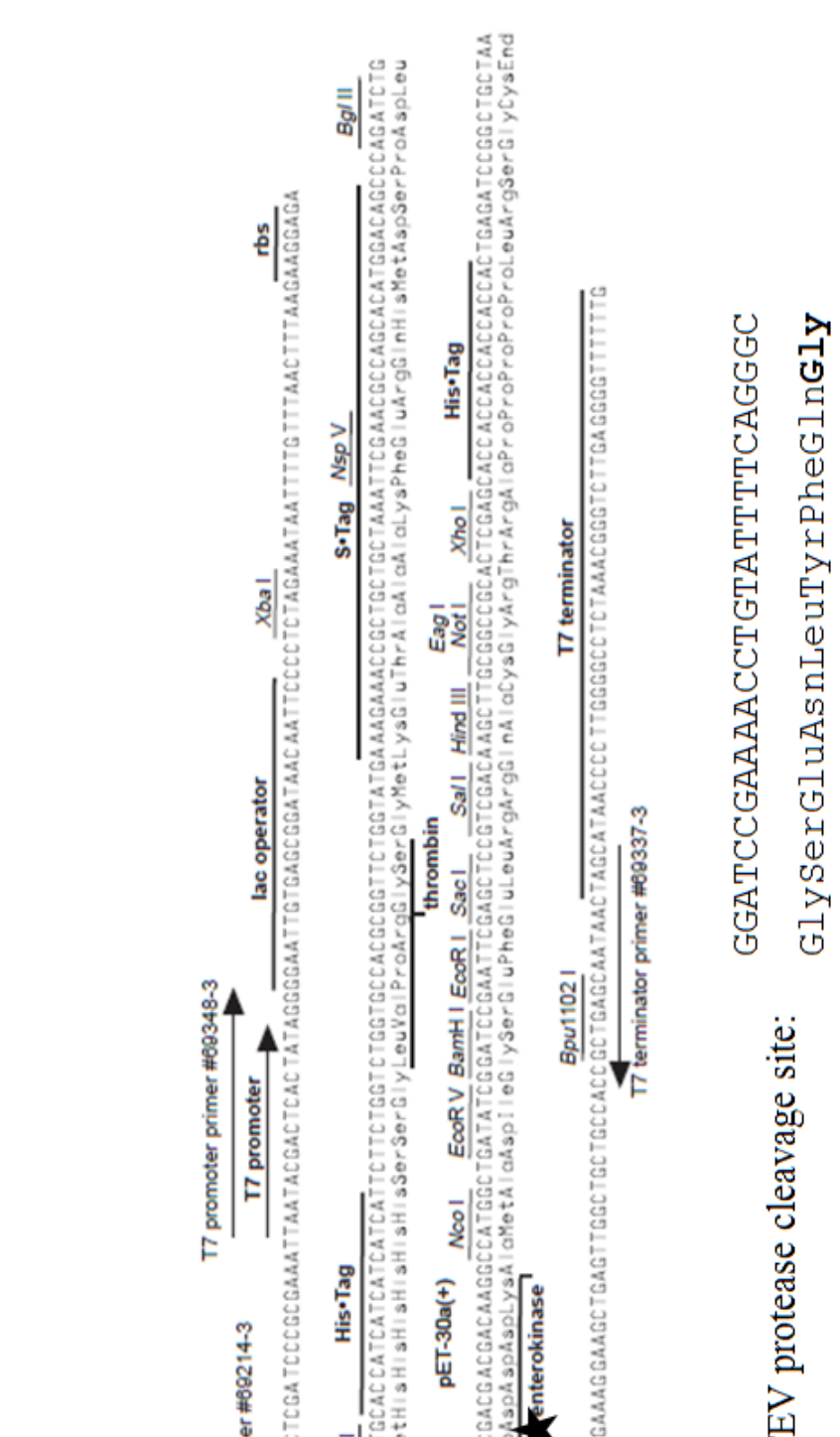


Figure A.1: Multiple-cloning site region of pET30a. T7 promoter and terminator, restriction enzyme sites, purification tags are shown. The red star indicates enterokinase cleavage site which was substituted with TEV protease recognition site (shown below). The glycine residue in bold is the amino acid retained within the recombinant protein sequences after cleavage.

Appendix B

Multiple sequence alignments

Applied colour	Residue at position	{Threshold, residue group}
Blue	A, F, I, L, M, V, W,	{+60%, WLVIMAFCHP}
	C	{+60%, WLVIMAFCHP}
Cyan	H, Y	{+60%, WLVIMAFCHP}, {+85%, W, Y, A, C, P, Q, F, H, I, L, M, V}
Green	N	{+50%, N}, {+85%, N, Y}
	Q	{+60%, KR}, {+50%, QE}, {+85%, Q, E, K, R}
	S, T	{+60%, WLVIMAFCHP}, {+50%, TS}, {+85%, S, T}
Magenta	D	{+60%, KR}, {+85%, K, R, Q}, {+50%, E, D}
	E	{+60%, KR}, {+50%, Q, E}, {+85%, E, Q, D}
Orange	G	{+0%, G}
Pink	C	{100%, C}
Red	K, R	{+60%, KR}, {+80%, K, R, Q}
Yellow	P	{+0%, P}

Table B.1: ClustalX default colouring scheme used by Jalview 2. Residues are coloured respectively according to the specified criteria. Threshold - minimum percentage presence of a specific amino acid or group. Commas separate residue groups. Source: http://www.jalview.org/help/html/colourSchemes/clustal.html.

Abbreviations used in Figures B.5 and B.5:
OCHIN - *Ochrobactrum intermedium;* BRUC - *Brucella* sp.; AGRVS - *Agrobacterium vitis;* XANC - *Xanthomonas campestris;* RHIME - *Rhizobium meliloti;* CHLT3 - *Chloroherpeton thalassium;* Desulfovib_alk - *Desulfovibrio vulgaris;* PHEZH - *Phenylobacterium zucineum;* RHILV - *Rhizobium leguminosarum;* PSEST - *Pseudomonas stutzeri;* RAMTT - *Ramlibacter tataouinensis;* Gemmata_obs - *Gemmata obscuriglobus;* KORVE - *Koribacter versatilis;* NITRD - *Nitrospira defluvii;* CHTFL - *Chthoniobacter flavus;* LEPIC - *Leptospira interrogans;* CHIPD - *Chitinophaga pinensis;* CALAB - *Caldithrix abyssi;* EMTOG - *Emticicia oligotrophica;* HALUD - *Halorhabdus utahensis;* HALJB - *Halalkalicoccus jeotgali;* RHOE1 - *Rhodococcus equi;* CYAP4 - *Cyanothece* sp.; FISCH - *Fischerella* sp.; NTRLA - *Nitrolancetus hollandicus;* ECOLI - *Escherichia coli;* PSEAE - *Pseudomonas aeruginosa;* AGRT - *Agrobacterium tumefaciens;* BACSU - *Bacillus subtilis;* PSESY - *Pseudomonas syringae.*

Figure B.1: Part of the multiple sequence alignment of all predicted HisKA_3 domains showing the region containing the invariant His. The HisKA_3 domain from an uncharacterised *Bacillus cereus* HK (UniProt entry B3ZQM2) was included to determine the position of the invariant His residue. Phosphorylatable histidine is indicated by the black arrowhead.

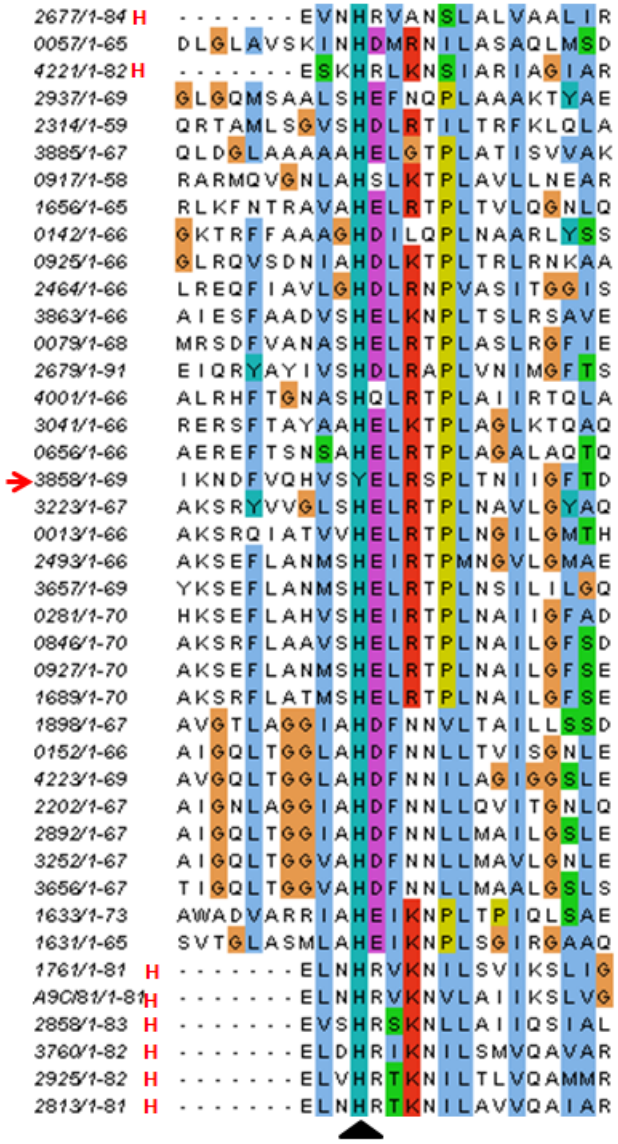


Figure B.2: Part of the multiple sequence alignment of all predicted HisKA and HWE_HK domains showing the region containing the conserved histidine residue. The HWE_HK domain of *Agrobacterium tumefaciens* BphP2 (UniProt entry A9CI81) was included to determine the localisation of the invariant His in HWE_HK domains. Phosphorylatable histidine is indicated by the black arrowhead.

Figure B.3: Part of the multiple sequence alignment of all predicted receiver domains showing the regions containing the conserved Asp and other active site residues (Asp/Glu and Lys). Gene ID followed by "recX" indicates receiver domains from hybrid HKs or RRs with more than 1 receiver domain; the receiver domains are numbered starting from N terminus. Phosphorylatable aspartate is indicated by black arrowhead, the other active site residues by red arrowheads.

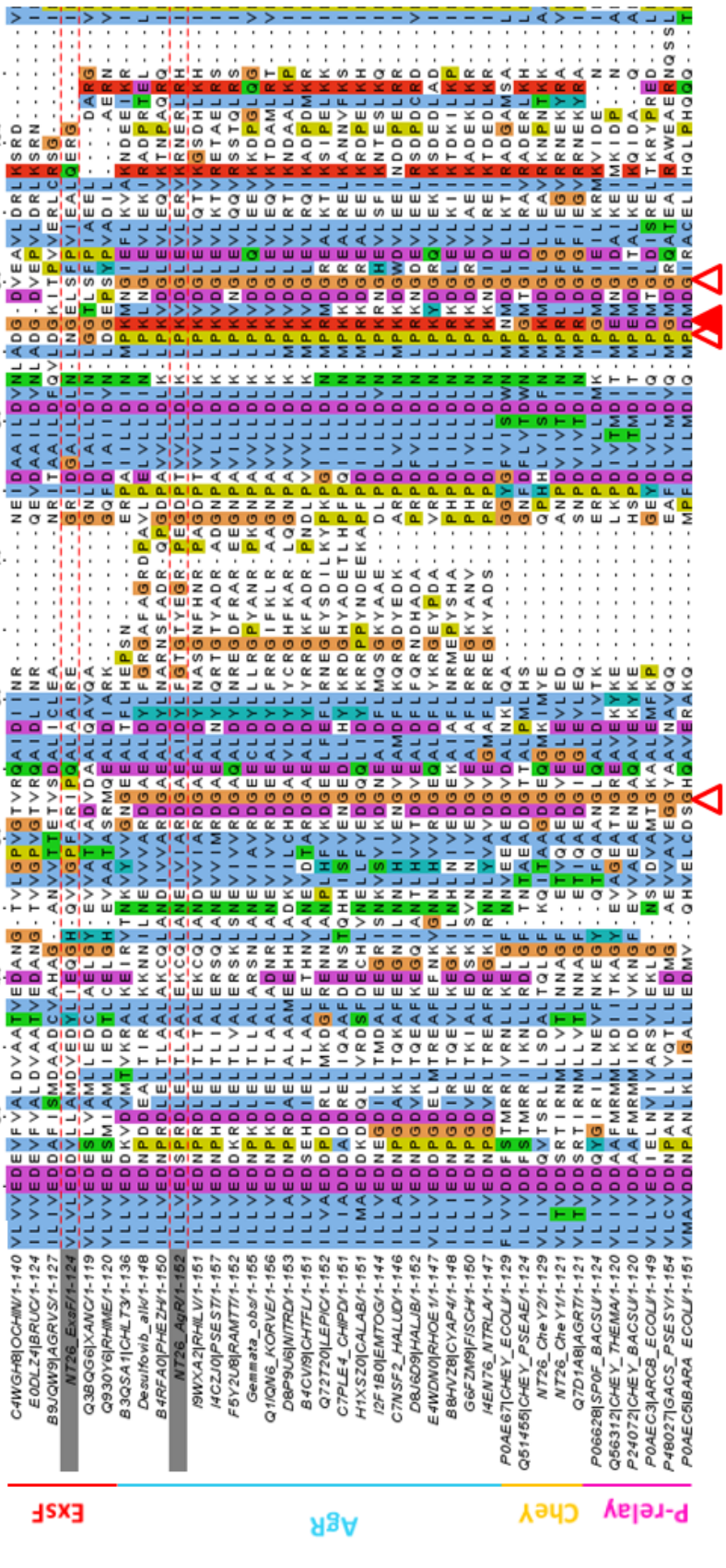


Figure B.4: Multiple sequence alignment of ExsF, AgR, their homologues, selected CheYs and receiver domains involved in phosphorelay (part 1). NT-26 AgR and ExsF are highlighted. Stars denote active site and Y-T coupling residues, filled star - phosphorylatable aspartate. Signature amino acid positions of ExsF are in red, of AgR - in cyan. Empty arrowheads signify absence of a conserved residue, filled - presence of a residue unique to the protein. Cyan lines denote AgR sequence extensions. P-relay - phosphorelay sequences. Codes represent UniProt entries, species codes are given earlier in Appendix C.

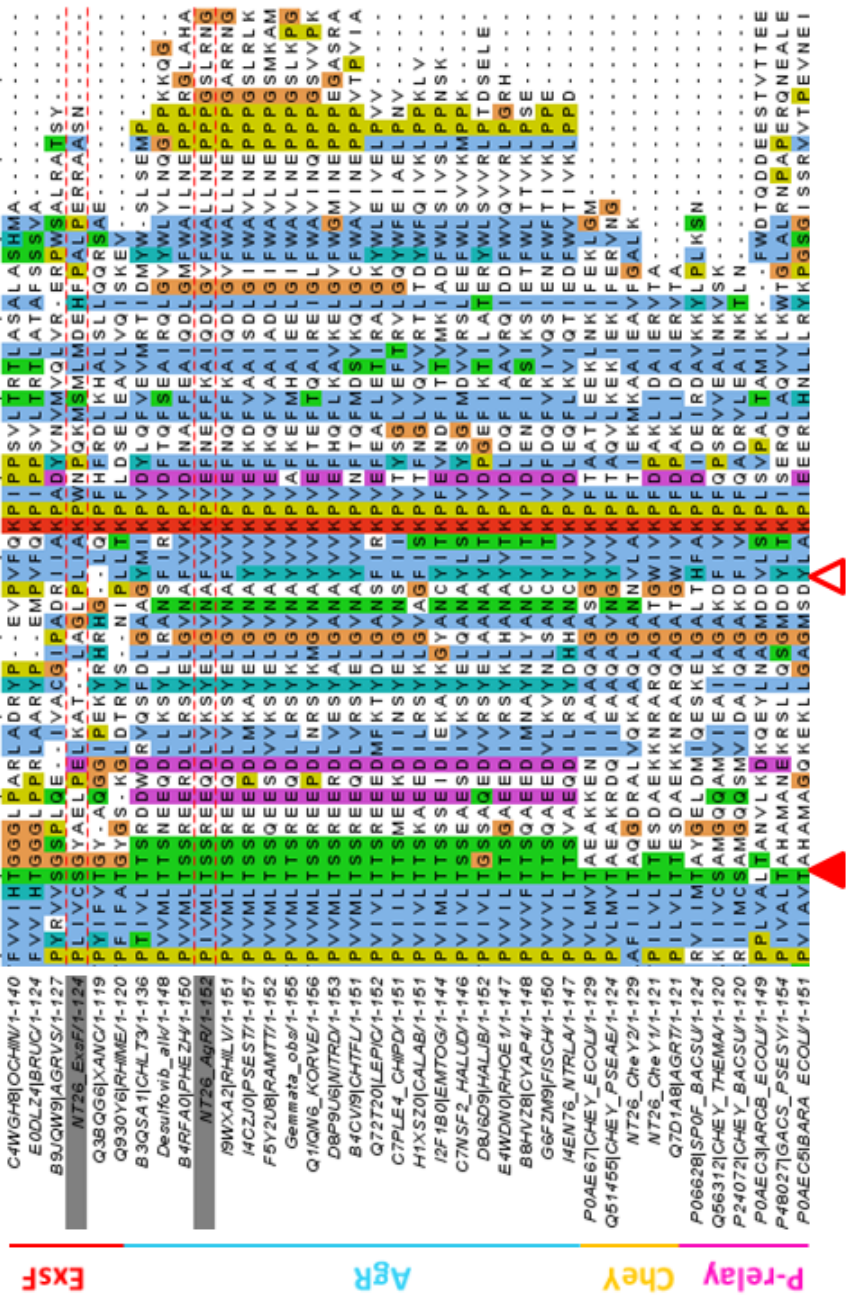


Figure B.5: Multiple sequence alignment of ExsF, AgR, their homologues, selected CheYs and receiver domains involved in phosphorelay (part 2).

Appendix C

LC-MS analysis of WT ExsG protein bands

Start	End	Sequence	Modifications
1	10	(-)GAMTVHNQLR(Q)	
1	10	(-)GAMTVHNQLR(Q)	Oxidation M (3)
1	10	(-)GAMTVHNQLR(Q)	Acetyl N-TERM (1)
14	28	(R)ILVLEDSHLDAELIR(E)	
29	41	(R)EHLETLKPVPEIR(Y)	
106	114	(R)QGATDYVLK(Q)	
106	116	(R)QGATDYVLKQR(L)	
117	126	(R)LIRLPAAVER(A)	
120	126	(R)LPAAVER(A)	
143	149	(R)QKELLVR(E)	
145	149	(K)ELLVR(E)	
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (7), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (5), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (5)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (5)
169	186	(R)TVKTSSSLDDYVENLLGR(L)	
172	186	(K)TSSSLDDYVENLLGR(L)	
189	212	(R)ALADAHALLFETNWNEAELLEVVR(R)	
225	238	(R)VRIDAGPLTLLTPK(A)	
227	238	(R)IDAGPLTLLTPK(A)	
239	256	(K)AALALSMVLNELATNAVK(H)	Oxidation M (7)
239	256	(K)AALALSMVLNELATNAVK(H)	
257	276	(K)HGALSGEHGTVHVSWQTDSK(G)	
277	282	(K)GVLTLR(W)	
283	292	(R)WQESDGPAVR(I)	
293	306	(R)IPEEAGFGMTLIER(S)	Oxidation M (9)
293	306	(R)IPEEAGFGMTLIER(S)	
310	317	(R)YELNGEVR(L)	

Table C.1: List of peptides and covalent modifications identified in the top protein band. The peptide containing the acetylated residue is highlighted.

Start	End	Sequence	Modifications
1	10	(-)GAMTVHNQLR(Q)	
1	10	(-)GAMTVHNQLR(Q)	Oxidation M (3)
14	28	(R)ILVLEDSHLDAELIR(E)	
14	41	(R)ILVLEDSHLDAELIREHLETLKPVPEIR(Y)	
29	41	(R)EHLETLKPVPEIR(Y)	
81	105	(R)DVAPEVPFIFVSGILGEEVAIESFR(Q)	
106	114	(R)QGATDYVLK(Q)	
106	116	(R)QGATDYVLKQR(L)	
117	126	(R)LIRLPAAVER(A)	
120	126	(R)LPAAVER(A)	
143	149	(R)QKELLVR(E)	
145	149	(K)ELLVR(E)	
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (7), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (7), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5)
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (5)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (5)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (5), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Phosphoryl STY (8)
169	186	(R)TVKTSSSLDDYVENLLGR(L)	
172	186	(K)TSSSLDDYVENLLGR(L)	
189	212	(R)ALADAHALLFETNWNEAELLEVVR(R)	
189	213	(R)ALADAHALLFETNWNEAELLEVVRR(T)	
213	220	(R)RTLAPQDR(G)	
214	220	(R)TLAPQDR(G)	
225	238	(R)VRIDAGPLTLLTPK(A)	
227	238	(R)IDAGPLTLLTPK(A)	
227	256	(R)IDAGPLTLLTPKAALALSMVLNELATNAVK(H)	Oxidation M (19)
239	256	(K)AALALSMVLNELATNAVK(H)	
239	256	(K)AALALSMVLNELATNAVK(H)	Oxidation M (7)
257	276	(K)HGALSGEHGTVHVSWQTDSK(G)	
277	282	(K)GVLTLR(W)	
283	292	(R)WQESDGPAVR(I)	
293	306	(R)IPEEAGFGMTLIER(S)	Oxidation M (9)
293	306	(R)IPEEAGFGMTLIER(S)	
310	317	(R)YELNGEVR(L)	
318	338	(R)LEYLETGLGCEIGIPLADLSG(-)	Carbamidomethyl C (10)

Table C.2: List of peptides and covalent modifications identified in the middle protein band. The peptide containing phosphorylated Ser163 is highlighted.

Start	End	Sequence	Modifications
1	10	(-)GAMTVHNQLR(Q)	Oxidation M (3)
1	10	(-)GAMTVHNQLR(Q)	
14	28	(R)ILVLEDSHLDAELIR(E)	
29	41	(R)EHLETLKPVPEIR(Y)	
81	105	(R)DVAPEVPFIFVSGILGEEVAIESFR(Q)	
106	114	(R)QGATDYVLK(Q)	
106	116	(R)QGATDYVLKQR(L)	
117	126	(R)LIRLPAAVER(A)	
120	126	(R)LPAAVER(A)	
143	149	(R)QKELLVR(E)	
145	149	(K)ELLVR(E)	
155	167	(R)VKNTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (9)
155	167	(R)VKNTMAMVMSLVR(R)	
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3), Oxidation M (5), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (5), Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (7)
157	167	(K)NTMAMVMSLVR(R)	
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (5)
157	167	(K)NTMAMVMSLVR(R)	Oxidation M (3)
169	186	(R)TVKTSSSLDDYVENLLGR(L)	
172	186	(K)TSSSLDDYVENLLGR(L)	
189	212	(R)ALADAHALLFETNWNEAELLEVVR(R)	
213	220	(R)RTLAPQDR(G)	
214	220	(R)TLAPQDR(G)	
225	238	(R)VRIDAGPLTLLTPK(A)	
227	238	(R)IDAGPLTLLTPK(A)	
239	256	(K)AALALSMVLNELATNAVK(H)	Oxidation M (7)
239	256	(K)AALALSMVLNELATNAVK(H)	
257	276	(K)HGALSGEHGTVHVSWQTDSK(G)	
277	282	(K)GVLTLR(W)	
283	292	(R)WQESDGPAVR(I)	
293	306	(R)IPEEAGFGMTLIER(S)	Oxidation M (9)
293	306	(R)IPEEAGFGMTLIER(S)	
310	317	(R)YELNGEVR(L)	
318	338	(R)LEYLETGLGCEIGIPLADLSG(-)	Carbamidomethyl C (10)

Table C.3: List of peptides and covalent modifications identified in the bottom (main) protein band.

Appendix D

Outline of the biophysical techniques

D.1 Phosphoprotein affinity PAGE

To separate phosphorylated and unphosphorylated protein forms, Phos-tag^{TM} acrylamide can be employed. The moiety of the Phos-tag^{TM} molecule binds to phosphoryl groups thus retarding the migration of phosphorylated proteins through SDS polyacrylamide gel. Singly or multiply phosphorylated proteins can be separated and visualised by Coomassie-based staining.

D.2 X-ray crystallography

X-ray crystallography can provide high resolution structures of proteins and other macromolecules, and has so far been the most efficient method for protein structure determination. In principle, X-rays are shone at a highly ordered array of molecules (a crystal) from different angles, and become scattered when they encounter electrons on their paths. Due to the symmetrical arrangement of the molecules within a crystal, the diffracted waves interfere with one another forming a unique diffraction pattern, from which the three-dimensional structure of the molecule can be inferred.
Crystallisation requires the state of supersaturation. In a typical crystallisation procedure, a drop containing the protein (highly concentrated) mixed with a precipitating agent solution is placed over (hanging drop) or next to (sitting drop) a reservoir containing the solution of precipitant. As the concentration of precipitant is higher in the reservoir, the water from the drop will slowly diffuse towards the reservoir, effectively increasing the concentration of the protein within the drop. Once the solubility limit of the protein is reached, it starts precipitating, and may form crystals if all other conditions like pH and salt concentration are right. Usually a few hundred of differ-

ent conditions need to be tested to find optimal ones which drive crystallisation. In addition, some crystals may need further refinement of conditions to increase their internal order and size, both of which can affect diffraction and thus on the quality of the acquired data.

D.3 Analytical ultracentrifugation (AUC)

Proteins and other biological macromolecules are usually too small to be affected by gravity as this force at the microscopic level is significantly outweighed by forces resulting from random motion of surrounding molecules. Centrifugation, i.e. rotating molecules in a cell at high speed, increases the effective gravitational potential energy by several orders of magnitude, allowing it to overcome the thermal energy and thus lead to sedimentation. AUC can be applied to studying both single species and a mixture of different macromolecules. The data can be acquired using interference and/or absorption optics systems - the scans represent concentration gradients formed along the cell. Given that the distance the sedimenting proteins travel is not infinite (an enclosed compartment), diffusion against the direction of the centrifugal force occurs when the concentration gradient becomes steeper. Therefore both sedimentation and diffusion affect the transport process of the macromolecule during centrifugation.

D.3.1 Sedimentation velocity

In sedimentation velocity experiments, a solution of one or several molecular species in a sector-shaped cell is subject to high speed centrifugation and the radial sedimentation of the molecules present is monitored. Larger and/or more compact molecules sediment faster, while smaller and/or ellipsoid ones migrate slower. Sedimentation profiles can provide information on homogeneity of the species present, their oligomeric state, molecular mass and shape of the molecule. Sedimentation process is described by the Lamm equation to which only analytical solutions exist, but numerical approximations can be found using programmes such as SEDFIT [156]. Both non-interacting and self-associating systems can be studied. SV uses high rotor speed and is a good method for determining the number of species present and their sedimentation coefficients, as well as the shape of the molecules. However, the molar mass distribution is extremely sensitive to errors in frictional coefficient estimation, hence the mass obtained for each species can significantly deviate from the actual mass.

D.3.2 Sedimentation equilibrium

Centrifugation at low rotor speeds creates a concentration gradient over a long period of time (usually over 24 hours). With increasing gradient, sedimentation starts being counteracted by back diffusion, and at some point the two forces balance each other, stabilising the gradient. This means that the flux of molecules sedimenting is equal to the flux of those diffusing away from the bottom of the cell. Such equilibrium state, once reached, can remain constant, provided the speed and composition of the sample do not change over time. The time needed for equilibration depends on the size, shape and heterogeneity of the sample, but concentration distribution at equilibrium is dependent solely on the molecular mass - hence the technique is much more accurate in estimating the mass of species present than SV experiments.

D.3.3 Data acquisition types

Interference data: in Rayleigh optical system a split beam of light is passed through the cell sectors containing the sample (protein solution) and the reference (solvent only), and then merged, with the waves forming an interference pattern (fringes). Differences in refractive index between corresponding points in the buffer and sample sectors result in shifting the fringes of the pattern, thus illustrating the concentration gradient along the sector.
280 nm absorbance data: absorption optics system measures the radial-dependent concentration gradient of a protein solution by recording the absorbance of light at 280 nm at each point along the sector. This method is based on the proportional relationship between optical density of a solution and solute concentration.

D.4 Mass spectrometry (MS) and ion mobility spectrometry (IMS)

D.4.1 Electrospray ionisation time-of-flight (ESI-ToF) MS

ESI-MS is a soft ionisation technique able to accurately determine the native mass of molecular complexes, their oligomeric states and oligomer forming units. In ESI the sample is present in an aqueous solution and is sprayed through a capillary needle. The capillary is held at a high electrical potential with respect to the entrance of the mass

spectrometer. Charged droplets are released from the end of the capillary. Consecutive Coulombic explosions produce multiply charged daughter droplets; further evaporation drives the formation of gas-phase ions containing multiply charged single molecules in question, whose velocities in the electric field applied differ depending on their mass-to-charge ratio. The actual mass of the molecule or complex studied can be determined from the peaks on the spectrum, which differ from one another by 1 charge.

D.4.2 Ion mobility spectrometry-mass spectrometry (IMS-MS)

Coupling ESI MS with IMS can provide information on conformational changes or stability of proteins in their native state or under stress. The ions generated in ESI MS are introduced into an IM chamber where they are subjected to counteracting flow of inert gas and an electric field. Depending on the collision cross-section, reflecting the shape, charge and mass of an ion, different ions will take different lengths of time to reach the detector. Ions with larger collision cross-sections exhibit longer drift times.

D.4.3 Liquid chromatography (LC)-tandem MS

This method is frequently used in proteomics, protein identification and determination of post-translational modifications. The proteins of interest are subjected to digestion which produces a number of peptides. Following LC chromatography (separation based on hydrophobicity), peptides enter the MS chamber and data are continually acquired. In the data-independent acquisition mode (LC-MS^E), energy in the collision cell is rapidly switched from low to high collision energy. When the energy is low, MS data is acquired for the intact peptide mass. When the energy is high, the peptides fragment about the peptide backbone in a sequence specific way, and MS/MS data is acquired. During data processing, fragment data (MS/MS) is assigned to a specific intact peptide mass (MS) based on a number of features including retention time, accurate mass, predicted fragmentation etc. This data is then search against a protein database. The database is *in-silico* digested to calculate all the theoretical peptides that could be present. From the MS and MS/MS data detected peptides are calculated the proteins present prior to digestion are inferred.

D.5 Isothermal titration calorimetry (ITC)

This technique can be used to determine the binding constant and thermodynamic parameters of protein-protein or protein-ligand interactions. In the latter case, the protein solution is contained in a cell kept at a constant voltage difference with respect to a reference cell. The ligand, usually present at 10-fold higher concentration than the protein, is introduced into the cell by several small-volume injections with constant stirring. Evolution or absorption of heat resulting from protein-ligand binding affects the voltage difference between the two cells and the amount of power that needs to be added or subtracted to recover the original difference can be used to calculate the binding constant, enthalpy, entropy, heat capacity, free energy and stoichiometry of the reaction.

D.6 Electron microscopy (EM)

EM allows low-resolution (5 - 20 Å) visualisation of proteins and protein complexes with molecular weight greater than 150 kDa. The experiment is performed under vacuum with an electron beam falling onto the sample positioned on a grid. In negative staining the stain fills in the space not occupied by biological material and is opaque to the electron beam thus appearing black. The biological sample is transparent but the edges and features are delineated by the stain particles and depending on the stain used (grain size), different contrast can be obtained. Uranyl acetate has the additional property of fixing the protein, but its drawback is the low pH (4-6) and coarse grain size (about 2 nm). Tungsten-containing stains often provide lower contrast but can be used at physiological pH, and offer a finer grain (about 1 nm).